

GenCore version 5.1.7  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2006, 09:07:29 ; Search time 42 Seconds  
 (without alignments)  
 1216.455 Million cell updates/sec

Title: US-10-721-553-2  
 Perfect score: 2764  
 Sequence: 1 MAPTIQTQAQREDGHRPNSH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : PIR\_80:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	645	23.3	425	2	T20261	hypothetical prote
2	335.5	12.1	547	2	C96828	unknown protein F1
3	280.5	10.1	457	2	T50233	probable DNA-direc
4	253	9.2	445	2	S44541	hypothetical prote
5	245.5	8.9	784	2	PN0009	neurofilament trip
6	241	8.7	792	2	T42963	hypothetical prote
7	236	8.5	590	2	A40437	glutamic acid-rich
8	234	8.5	953	2	C89824	hypothetical prote
9	230.5	8.3	1385	2	D89824	hypothetical prote
10	230	8.3	877	2	F90070	Clumping factor B
11	225.5	8.2	933	2	S41539	fibrinogen-binding
12	224	8.1	797	2	A36811	hypothetical prote
13	223.5	8.1	1315	2	T28679	fibrinogen-binding

14	221	8.0	1192	2	A71623	probable secreted
15	220	8.0	334	2	A54138	acidic repetitive
16	218	7.9	406	2	S38170	SRP40 protein - ye
17	218	7.9	1166	2	T28680	fibrinogen-binding
18	216	7.8	913	2	T52485	neurofilament prot
19	216	7.8	1110	2	I51116	NF-180 - sea lamp
20	214.5	7.8	1141	2	E89824	hypothetical prote
21	214	7.7	727	2	JC5113	ribosomal transcri
22	214	7.7	2364	2	A56577	microtubule-associ
23	213.5	7.7	678	2	A54514	glutamic acid-rich
24	212.5	7.7	1092	2	T30214	fibrinogen-binding
25	211.5	7.7	665	2	B71609	hypothetical prote
26	211	7.6	989	2	D89852	fibrinogen-binding
27	208.5	7.5	798	2	I50479	neurofilament medi
28	208	7.5	765	2	S22314	transcription fact
29	203.5	7.4	606	2	S70358	centromere protein
30	202.5	7.3	764	2	JC5112	ribosomal transcri
31	202	7.3	727	2	S18193	autoantigen NOR-90
32	202	7.3	1526	2	A45605	mature-parasite-in
33	200.5	7.3	407	1	EDBEQ3	immediate-early pr
34	199.5	7.2	599	2	S18735	centromere protein
35	198.5	7.2	524	2	S35551	transcription fact
36	198	7.2	727	2	B40439	UBF transcription
37	195.5	7.1	500	2	S55785	nucleolar protein
38	195	7.1	764	2	S09318	transcription fact
39	194	7.0	409	2	E86336	hypothetical prote
40	193	7.0	1173	2	T42719	TPR-containing/SH2
41	192.5	7.0	2464	1	QRMSP1	microtubule-associ
42	192	6.9	1178	2	S78475	mannosylphosphoryl
43	191.5	6.9	764	2	A40439	UBF transcription
44	191.5	6.9	852	2	A34373	histidine-rich cal
45	191	6.9	598	2	B40713	cylicin I - human

#### ALIGNMENTS

##### RESULT 1:

T20261

hypothetical protein C55A6.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T20261

R;Kershaw, J.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19243

A;Accession: T20261

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-425 <WIL>

A;Cross-references: UNIPROT:P90783; UNIPARC:UPI00000748D2; EMBL:Z81051;

PIDN:CAB02869.1; GSPDB:GN00023; CESP:C55A6.9

A;Experimental source: clone C55A6

C;Genetics:

A;Gene: CESP:C55A6.9

A;Map position: 5

A;Introns: 14/2; 48/2; 90/3; 177/3; 381/1

Query Match 23.3%; Score 645; DB 2; Length 425;  
 Best Local Similarity 33.1%; Pred. No. 1.5e-26;  
 Matches 146; Conservative 96; Mismatches 165; Indels 34; Gaps 10;

```

Qy      24 PERSGVVCrvKYCNslPDlPFDPKFITYPF-DQnRFVQYKATSLEKQHKHDLlTEPDlGV 82
      | :  : : : : | : | : | | | | | | | | | | : : : : | : | : : | : | :
Db      14 PRKVDFMLKPRFTNTVPDVPFDaKfMTCpFVPLGRfVEfQpAAIYRDYKHAVICDDDMGL 73

Qy      83 TIDLINPDtYRIDP-NVLLDPADEKLLEEEIQAptSSKRSQQHAKVVPWMrKtEYISTEF 141
      : | | : | | | : | | | | : | | | : | | | : | | | : | | | : | | | :
Db      74 NVDLIDLKKYDEdPIETEIdEKDNIlLEDdGAakLIaKRSQQHskLVpWMrKtEYISTEF 133

Qy     142 NRYGISNEKPEVKIGVSVKQQfTEEEIYKDRDSQITaIEKtFEDaQKSISQhYSKPRVTP 201
      | | : | : : | | : | : | : | : | : | | | | | | | | : | : : | | | |
Db     134 NRFGVTADRQETKLGYNLKKnQQVEDMYRDKQSQIDaINKtFEDVRKpVKEhYSKKGvKA 193

Qy     202 VEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEMMSQAMIRGMMDEEGNQfVAYF 261
      | | | | | | | | : | | | | | : : : : | : | : | | | |
Db     194 VEESFVFPDFdHWKHLfAHVQfDGDtITTEfEEEDERQQARESSVIKAMEfEDQKfAAVF 253

Qy     262 LPVEETLKKRKRdQEEEMdYAPDDVYDYKIAREYNWNVNKNKASKGYEENYffIFREGdGV 321
      : | | | | | : | | : : | | : : : : : : : | : : :
Db     254 VPTIGCLTSFMDDLELERPFDEdMKYEFLLSREYtFKMEHLPPR--DRDVFIMYHRNNVF 311

Qy     322 YYNELETRVRLSKRRaKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLENHE-PEEEE 380
      | | | : | : : | : : : | : : | | | : | : | : | : | :
Db     312 QYNEVDCNVKMT-RKPKMALSRKSKLTLTYR----NPSELEQKDMNKREAElyEQPKTRK 366

Qy     381 EEEME--TEEKEAGGSDEEQEKGSsSEKEGSedeHSGSEsEREeGDRDEASDKSGSGEDE 438
      : | : | | : | | | | | | | : | : : : | | | |
Db     367 QEILEKIQEKKEEGGD-----SSDQSSDSDDDKPQKSR-----SDSSSDV 406

Qy     439 SSEDEARAARDKEEIfGSDAD 459
      | | : : | | | | : |
Db     407 SSDDD--SPRKKEPTVDSdSD 425
  
```

# RESULT 2

C96828

unknown protein F19K16.29 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C;Accession: C96828

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;

Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;

Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;

Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;

Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;

Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;

Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-

Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,

S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,

M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: C96828

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-547 <STO>

A;Cross-references: UNIPROT:Q9CA82; UNIPARC:UPI00000A4648; GB:AE005173;

NID:g6453869; PIDN:AAF09053.1; GSPDB:GN00141

C;Genetics:

A;Gene: F19K16.29

A;Map position: 1

Query Match 12.1%; Score 335.5; DB 2; Length 547;

Best Local Similarity 24.8%; Pred. No. 2.1e-10;

Matches 121; Conservative 69; Mismatches 136; Indels 161; Gaps 20;

```
Qy      6 QTQAQRED-GHR---PNSHRT-----LPERSGVVCRVKYCN 37
      : : ||:|  ||    |||::          | : : :||::| |
Db      143 ELEKQRQDEKHRQQMKNSHKSQMPKGHTEEKKPTPLLTDRVENRLKKPTTFICKLKFRN 202

Qy      38 SLPDIPFDPKFIT-----YPFDQNRVQYKATSLEKQHKHDLLETPDLGVTIDLINPDT 91
      |||    | :|    : ||  || :|  ||||  | :  ||||: :||::
Db      203 ELPDPSAQLKLMTIKRDKDHYFDPTRFKYTITSLEKLWKPKIFVEPDLGIPDLDDLDSV 262

Qy      92 YRIDPNVL--LDPADKLLLEEIIQAPTSSKRSQQHAKVVP-----WMRKTEYISTEFNR 143
      |  | |  | | ||:| | :|  | :|  |  |  | : ||:| | |
Db      263 YN-PPKVKAPLAPEDEELLRDD-DAVTPIKKDGIRRKERPTDKGMSWLVTQYISS---- 316

Qy      144 YGISNEKPEVKIGVSVKQQTTEEEI-----YKDRDSQITAIEKTFEDAQK 188
      |:| |      | :|  ||::          :|: ||  || :| | |
Db      317 --INNE-----SARQSLTEKQAKELREMKGGINILHNLNNRERQIKDIEASFE-ACK 365

Qy      189 SISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDS-----DPAPKDTSGAA 236
      |  | :  : |||:|: | | :      ||      ||: :|  :
Db      366 SRPVHATNKNLQPVEVLPLLPHYFDRYDEQFVVANFDGAPIADSEFFGKLDPSIRDAHESR 425

Qy      237 ALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYDYKIAREYN 296
      |:  :  :: |      :|:| | :|  | |  | : ||: |      |||
Db      426 AI--LKSYYVAGSDTANPEKFLAYMVPSPDELSDIHDENEEISYT-----WVREYL 475

Qy      297 WNVKNKASKGYEENYFFIFREGDGYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDM 356
      |:|: | :      | |      :|| : ||  |
Db      476 WDVQPNAN-----DPGTYL-----VSFDNGTASYLVYSSR-- 505

Qy      357 NEKELEAQEARKAQLN-----HEPEEEEEEMETEEKEAGGSDEEKEGSSSEKE 407
      : |  :: :||:      |||::          ::  :
Db      506 ----VGASSSKMRRLEDEGGLGRSWKHEPEQD-----ANQYSD 539

Qy      408 GSEDEHS 414
      |:|:|:|
Db      540 GNEDDYS 546
```

T50233

C;Species: Schizosaccharomyces pombe

C;Accession: T50233

submitted to the EMBL Data Library, January 1999

A;Reference number: Z25048

A;Accession: T50233

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-457 <CAD>

A:Cross-references: UNIPROT:O9US06; UNIPARC:UPI000006A70C; EMBL:AL136235;

PIDN:CAB65804.1; GSPDB:GN00066; SPDB:SPAC664.03

A;Experimental source: strain 972h(-); cosmid c664

C; Genetics:

A;Gene: SPDB:SPAC664.03

```
A;Map position: 1
```

A; Introns: 1/3

Query Match 10.1%; Score 280.5; DB 2; Length 457;

Best Local Similarity 24.9%; Pred. No. 1.2e-07;

Matches 122; Conservative 94; Mismatches 192; Indels 81; Gaps 21;

Qy 26 RSGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQYK----ATSLEKQHKHDLLETPDLG 81

Db 6 RODYILRVRYHNPLPPPPFPPKLINIP---NPVKQYALPNFVSTLVQEKKIPIENDIELG 62

QY 82 VTIDL-----INPDYRID---PNVLLDPADEKLLLEEIIQAPTSSKRSQQHAKVVPWMR 132  
: : | | : : | | | | : : : : : : | : | : | :

Db 63 MPLDLAGITGFFEGDTSWMHSDLSSVNLDPIDRSLLK-----VAGGSGSTHLE-VPFLR 115

**QY**:  
133 KTEYISTEFNRYGISNEKPEVKIGSVKQQFTEE--EIYKD RDSQTITAIEKT FEDA QKSI 190  
| | : | | : : : | : : : | : || :|| :|| :

Db 116 RTEYISSEVAR--AASNRGNLRLTASTSKALAEQRGRSLREVPKQLEAINESFDVVQQPL 173

QŸ 191 SQ--HYSKPRVTPVEVMPVFDFKMWINPCAQVIFDSDPAPKDTSGAA----- 236

Db : 174 EOLKHPTKPD LKPVSAWNLLPNTSMAGIQHMLRVADDLSERSHSYSSLVNLQEGHNLTk 233

QY 237 ----ALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRDDQEEMDYAPDDVDYDKIA 292  
||| | ||| : :::: ||| : : :: | ||: :

Db 234 RHEVALFMPSSA-----EGEEFLSYLPSEETAET---EIQAKVNDASADVHEPFVY 281

Qy 293 REY-NWNVKNKASKGYEENYFFIFR-----EGDGVYYNELETRVRLSKRRAKAGVQSGT 345  
: |:: : | : | : | : | : | : | : |

Db : 282 NHFRNFDASMHVNSTGLEDLCLTFHTDKDHPEANQVLYTPIYARSTLKRRHVRAPVSLDA 341

```
QY      346 NALLVVKHRDMNEKE-LEAQEAR--KAQLENHEPEEEEEEMETEEKEAGGSDEEQEKG 402
          :  :  ||:|:|  | :  : ||  |  :  |||:|:| :  :  |  : ||
```

Db 342 VDGIELSLRDLNDEESLQLKRARYDTFGLGNIKDLEEEEEKLRSVE---GSLNEE----L 394

QY            403 SSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSSEDEARA-ARDKEEIFGSDADSE 461  
| |:::|                 ::| :|        ::|        :|| :| :|        ||    |:    |

Db 395 SEEKPAESREQLESAEQTNGVKPETQAQNMS----ASESQANSPAPPVEE--GNTQPSP 448  
QY 462 DDADSDDDED 470  
: :||  
Db 449 VEQLQNEED 457

RESULT 4

S44541

hypothetical protein YBR279w - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein YBR2016; Paf1 protein

C;Species: *Saccharomyces cerevisiae*

C;Date: 08-Jun-1994 #sequence\_revision 09-Sep-1994 #text\_change 09-Jul-2004

C;Accession: S44541; S46161; JC6088; PC6031; S39135

R;Holmstrom, K.; Brandt, T.; Kallesoe, T.

Yeast 10(Suppl.A), S47-S62, 1994

A;Title: The sequence of a 32420 bp segment located on the right arm of chromosome II from *Saccharomyces cerevisiae*.

A;Reference number: S44537; MUID:94378722; PMID:8091861

A;Accession: S44541

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-445 <HOL>

A;Cross-references: UNIPROT:P38351; UNIPARC:UPI0000053037; EMBL:X76053;

NID:g600025; PIDN:CAA53642.1; PID:g429124

R;Brandt, T.; Christiansen, C.; Holmstroem, K.; Kallesoe, T.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S46157

A;Accession: S46161

A;Molecule type: DNA

A;Residues: 1-445 <BRA>

A;Cross-references: UNIPARC:UPI0000053037; EMBL:Z36148; NID:g536721;

PIDN:CAA85243.1; PID:g536722; MIPS:YBR279w

R;Shi, X.; Finkelstein, A.; Wolf, A.J.; Wade, P.A.; Burton, Z.F.; Jaehning, J.A.

Mol. Cell. Biol. 16, 669-676, 1996

A;Title: Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription.

A;Reference number: JC6088; MUID:96140434; PMID:8552095

A;Accession: JC6088

A;Molecule type: DNA

A;Residues: 1-166,168-445 <SHI>

A;Cross-references: UNIPARC:UPI0000179A39

A;Experimental source: strain YJJ453

A;Accession: PC6031

A;Molecule type: DNA

A;Residues: 5-11;420-427 <SH2>

A;Cross-references: UNIPARC:UPI0000179A3A; UNIPARC:UPI0000179A3B

C;Comment: This factor is a highly charged nuclear protein, and acts as a cofactor important for transcriptional activation and repression from diverse promoters.

C;Genetics:

A;Gene: SGD:PAF1

A;Cross-references: SGD:S0000483; MIPS:YBR279w

A;Map position: 2R

A;Note: this gene is located at the right arm of chromosome II

C;Superfamily: *Saccharomyces cerevisiae* hypothetical protein YBR279w

C;Keywords: nucleus

F:25-49/Region: PEST sequence  
F:119-141/Region: nuclear location signal

Query Match 9.2%; Score 253; DB 2; Length 445;  
Best Local Similarity 22.1%; Pred. No. 3.2e-06;  
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;

```
Qy      23 LPERSGVVCRVKYCNSLPDIPFDPKFITYP-----FDQNRVQYKATSLEKQHK 71
      : :: : :|| ||| || : || : :: :| :|
Db      1 MSKKQEYIAPIKYQNSLPVPQLPPKLLVYPESPETNADSSQLINSLYIKTNVTNLIQQ-- 58

Qy      72 HDLLTEPDLGVTIDLI-----NPDYRIDPNVLLDPADEKLLEEEIQAPTSSKRS 121
      : ||| :||| : : | | || | | || : | : :
Db      59 -----DEDLGMPVDLMKFPGLLNKLDKLLYGFD-NVKLDDKDRILLRD----PRIDRLT 108

Qy     122 QQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQ----- 175
      : | :| :||| : : : | : : : | ||
Db     109 KTDISKVTFLLRTEYVSNTIAAHDNTSLKRKRRL-----DDGSDDENLDV 154

Qy     176 ---ITAIEKTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDT 232
      |: :| || | || | | | : : || || |
Db     155 NHIISRVEGTFNKTDK--WQHPVKKGVMKVKWDLDPD----TASMDQVYF-----ILKF 203

Qy     233 SGAAALEMMSQAMIR-GM---MDEEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYD 288
      |:|:| : : : | : : | : : : : : : : : | : || || : :
Db     204 MGSASLDTKEKKS LNTGIFRPVELEEDEWISMYATDHKDSAILENELEKGMDEMDDDSHE 263

Qy     289 ---YKIAREYNWNVKNKASKGYEENYFFIFREGDGV-YYNELETRVRLSKRRRAKAGVQSG 344
      || | : : :| | | | : ||| | : : | : || : :
Db     264 GKIYKFKRIRDYDMKQVAEKPMTE-LAIRLNDKDGIAYYKPLRSKIELRRRRVNDI IKP- 321

Qy     345 TNALLVVKH-----RDMNEKELEAQEARKAQLN-----HEPEEEEEEEEMETEEK 389
      || :| | : : || : : : : : | :| :| :| :|
Db     322 ----LVKEHDIDQLNVTLRNPSTKEANIRDKLRMKFDPINFATVDEEDEDEEQPEDVKK 377

Qy     390 EAGGSDEEQEKGSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARD 449
      |: | : : :|| | | :|| | | | : :| : :|| |||
Db     378 ESEG--DSKTEGSEQEGENEKDEEIKQEKENEQ-----DEENKQDENRAADT 422

Qy     450 KEEIFGSDADSEDDADSDDEDRGQAQ 475
      | ||| : : : : | :
Db     423 PET---SDAVHTEQKPEEEKETLQEE 445
```

RESULT 5

PN0009

neurofilament triplet M protein - Pacific electric ray (fragment)

C;Species: Torpedo californica (Pacific electric ray)

C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 09-Jul-2004

C;Accession: PN0009

R;Linial, M.; Scheller, R.H.

J: Neurochem. 54, 762-770, 1990

A;Title: A unique neurofilament from Torpedo electric lobe: sequence, expression, and localization analysis.

A;Reference number: PN0009; MUID:90155300; PMID:2106008

A;Accession: PN0009

A;Molecule type: mRNA

A;Residues: 1-784 <LIN>  
A;Cross-references: UNIPROT:Q7LZ90; UNIPARC:UPI00001774FE  
C;Comment: Neurofilaments are a subgroup of intermediate filaments which are expressed specifically in neuronal cells.  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; cytoskeleton; intermediate filament; nerve; phosphoprotein; tandem repeat  
F;1-52/Region: serine-rich  
F;53-84/Region: coil Ia  
F;98-194/Region: coil Ib  
F;217-367/Region: coil II  
F;400-597/Region: glutamic acid-rich  
F;598-674/Region: 6-residue repeats  
F;675-784/Domain: carboxyl-terminal #status predicted <CTD>  
F;616,622,628,634,640,646,652,658,670/Binding site: phosphate (Ser) (covalent)  
#status predicted

Query Match 8.9%; Score 245.5; DB 2; Length 784;  
Best Local Similarity 20.5%; Pred. No. 1.5e-05;  
Matches 122; Conservative 102; Mismatches 229; Indels 143; Gaps 23;

Qy	47	KFITYPFDQNR	FVQYKATSLEKQ	---	HKHDLLTEPDLG	-----	VTIDLINPDTYR	93
		:		:	:::		:	:
Db	62	RFAGY-IDKVHYLEQ	QNKELAEIQAH	RQKQVSHGQLG	DVYDQEIRELRS	IEQVNQEKAQ		120
Qy	94	ID-PNVLLD	-----	PADEKLLEEE	-----	IQAPTSSKRSQQ	HAKVV	128
			:		:		:	
Db	121	IQLDSVHLDDDFQ	RVGAFDEEALR	DEDPEATIRVL	KKETEDSVIQAG	DGEKKAQSLQ	DEV	180
Qy	129	PWMRKTEYISTEF	NRYGISNEKPEV	KIGVSVKQQFTE	EEIYKDRDSQIT	AIEKTFEDAQK		188
		::	:		:		:	:
Db	181	AFLR-----	NNHEEEV-ADL	FQAQIATQVTVEK	-KDFLKTDITS	SALKEIRS		224
Qy	189	SISQHYSKPRVTP	PVEVMPVFPDF	KMW-----	INPCAQVIFDSD	PAPKDTSGAA	ALEMM	241
		:		:		:	:	:
Db	225	QLEGHSAKNMQQ	ADE-----	WFKCRYDKLNEA	EAEMNKDAIRA	AREEIGEYRR	QLQ	274
Qy	242	SQAM-----	IRGM-----	MDEEGNQFVAY	FLPVEETLKKR	KRDQEEEM----	DYA	282
		:::	:	:		:		:
Db	275	SKSIELESVRST	KESLERQLTDI	EDRHNADVANY	QETVQQLENEL	RGTKWEMARH	LREY-	333
Qy	283	PDDVYDYKIARE	YNWNVKNKASK	GYEENYFFIFR	EGDGVYYN----	ELETRVRLSK	RRA	337
		:	:	:	:	:	:	:
Db	334	-QDLLNVKMALD	IEIAAYRKLE	GEESRYTF-SGT	GPSIPYRSPSR	PRLPKAVHKT	KEVP	391
Qy	338	KAGVQSGTNALL	VVKHRDMNEK	-----	ELEAQEARKAQ	LENHEPEEEEE	EMETEE-----	388
				:	:	:	:	:
Db	392	KVKVQHKFVEE	IIETKVKDEKA	EMGDI DLAE	AVEGGATMES	PEDKEEA	AEKVVEEAIVAT	451
Qy	389	-----	KEAGGSDEEQE	KGSSEKEGSE	DEHSGSESERE	EGDRDEASDK	SGSGEDESSED	442
				:	:			:
Db	452	VKAGVQAEPRGE	AESEAKEEED	EGVEEEKEE	-EADDEEKGE	EKDEEGEA	EDEAEGG	510
Qy	443	EARAARDKEE	IF-----	GSDADSEDDAD	SDDEDRGQAQ	GGSDNDS	SGSNGGGQR-	492
		:		:		:	:	:
Db	511	ESRVVEEKVEI	VKVEQSKAHP	GKDEVKEERKE	EEEEEEGE	ASGESDKE	STGGAINGSQEE	570

Qy 493 -----SRSHRSASPFPSGSEHSAQED----GSEAAASDSSEADSDS 530  
 Db 571 SKGKVEEKLTVTEKATEDKVSPREEKPQKEEQKDIEEKKEEAKSKDEAKSKDEA 626

RESULT 6:

T42963

hypothetical protein 48 - ateline herpesvirus 3 (strain 73)

C;Species: ateline herpesvirus 3

A;Variety: strain 73

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004

C;Accession: T42963

R;Albrecht, J.C.; Fleckenstein, B.

submitted to the EMBL Data Library, August 1998

A;Description: Primary structure of the herpesvirus ateles genome.

A;Reference number: Z22274

A;Accession: T42963

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-792 <ALB>

A;Cross-references: UNIPROT:Q9YTL7; UNIPARC:UPI000000EC1E3; EMBL:AF083424;

PIDN:AAC95573.1

A;Experimental source: strain 73

Query Match 8.7%; Score 241; DB 2; Length 792;

Best Local Similarity 21.2%; Pred. No. 2.6e-05;

Matches 94; Conservative 69; Mismatches 173; Indels 108; Gaps 14;

Qy 169 YKDRDSQITAIEKTFEDAQKSI-----SQHYSKPRVTPVEVMPVFP 209  
 Db 213 YQYMSSDLIAIEEALQSSYLSICGSTYPSYSKILELLTANMSKEHIRQKVNVD----- 266

Qy 210 DFKMWINPCA-QVIFSDPAPKDTSGAAALEMMSQAMIRGM-----MDEEGNQ 256  
 Db 267 ----FIKPSLHQMIRDTKKEPRQKTKTLMISILGS---RGIGLDLFRQTQDVLKFPSSDAK 319

Qy 257 FVAYFLP-----VEETLKKRRKDQEEEMDYAPDDVYDYKIAREYNWNVKNKASK---- 305  
 Db 320 FMAVSQPDNFNEKEVEFSMTGGKTDSEEDVT--APRKVGKNSLNRYLENLKNKRNKNNY 377

Qy 306 -GYEENYFFIFREGDG-----VYNELE-----TRVRLSKRRAKAGVQSGTNAL-- 348  
 Db 378 SGRNNKY-----KGDGANDKDKSIDKNESEGGDHSEINREKNRKRKKPNGFRVGDKEVGE 432

Qy 349 -----LVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEK 389  
 Db 433 EKSVKSGEGKKSEKDSEEEAEDKDEEENKKKGDGEDEDEDEDEDEDEDEDEDEDEDEDEDE 492

Qy 390 EAGGSDEEQEGKSSSE--KEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAA 447  
 Db 493 EEEEEDE 552

Qy 448 RDKEEIFGSDADSEDDADSDDEDRGQAQGGSDNDSDSGSGNGGGQSRSHRSASPFPSGS 507  
 Db 553 EEEEEDE 605

Qy 508 EHSAQEDGSEAAASDSSEADSDSD 531

Db 606 EKEDDEDDEEEEDDEEEEDDEE 629

RESULT 7

A40437

glutamic acid-rich protein, retinal - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 09-Jul-2004

C;Accession: A40437

R;Sugimoto, Y.; Yatsunami, K.; Tsujimoto, M.; Khorana, H.G.; Ichikawa, A.

Proc. Natl. Acad. Sci. U.S.A. 88, 3116-3119, 1991

A;Title: The amino acid sequence of a glutamic acid-rich protein from bovine retina as deduced from the cDNA sequence.

A;Reference number: A40437; MUID:91195303; PMID:2014230

A;Accession: A40437

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-590 <SUG>

A;Cross-references: UNIPROT:Q28181; UNIPARC:UPI000016C311; GB:M61185;

NID:g163077; PIDN:AAA30536.1; PID:g163078

Query Match 8.5%; Score 236; DB 2; Length 590;

Best Local Similarity 22.4%; Pred. No. 3.3e-05;

Matches 133; Conservative 80; Mismatches 221; Indels 160; Gaps 28;

```
Qy 21 RTLPERSGVVCRVKY--CNSLPDIPFDPKFITYPFDQNRVQYKATSLEKQHKHDLLEP 78
   | ||: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 7 RVLPPQPPGTPQKTKQEEEGTEPEPELEPKPETAPEE----TELEEVSLPPE-----EP 55

Qy 79 DLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTS---SKRSQQHAKVVPWMRK-- 133
   :| : : : | | : || | | : | : | : | : | : | : | : | : | : |
Db 56 CVGKEVAAVTLGPQGTQETALTPPT-----SLQAQVSVAPEAHSSPRGWLTWLRKGV 108

Qy 134 -----TEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE----- 180
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 109 EKVVPQPAHSSRPSQNIAGLESPPDQQAQILGQCGTGG--SDEPSEPSRAEDPGPGPW 166

Qy 181 --KTFE-DAQKSISQHYSKPRVT-----PVEVMPVF----- 208
   : || : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 167 LLRWFEQNLEKMLPQ---PPKISEGWRDEPTDAALGPEPPGPALEIKPMLQAQESPSLPA 223

Qy 209 -----PDFKMWINPCAQVIFDSDPAPKDTSGAAA-----LEM-MSQAMIRGMMDEEGNQF 257
   | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 224 PGPPEPEEEPIPEPQPTIQASSLPPQDSARLMAWILHRLEMALPQPVIRGKGGEQESD- 282

Qy 258 VAYFLPVE---ETLKKRKRQEEEMDYAPDDVYDYKIAREY--NWNVKNKASKGYEENYF 312
   || : : : || : | : | : | : | : | : | : | : | : | : | : | : |
Db 283 ----APVTCDVQTISILPGEQEE-----SHLILEEVDPHW----- 313

Qy 313 FIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKH-----RDMNEKELEAQEA 366
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 314 ----EEDEHQEGSTSTSPRTSE-AAPADEEKGE----VVEQTPRELPRIQEKEDEEEEEK 364

Qy 367 RKAQLENHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSE-DEHSGSESEREEGDR 425
   : | | : |||| : : | : | : | : | : | : | : | : | : | : | : |
Db 365 EDGEEEEEEGREKEEEEGEKEEEEE-GREKEEEEGEKKEEEEGREKEEEEGGEKEDEEGRE 423
```

```

Qy      426 DEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSD-----DEDRG 472
      | : | |:| | | | | | : | : : : : | : | :
Db      424 KEEEEGRGKEEEEKGKEEEEGRGKEEVEGREEEDEEEEQDHSVLLDSYLVLPQSEEDQS 483

Qy      473 QAQGGSDNDSDSGSGNGGQSRSHRSASPFPSGSEHSAQEDGSE-AAASDSSE 525
      : : : : | | | : : | | | : | | | | |
Db      484 E-----ESETQDQSEVGGAQTQGEVGGAQAL---SEESSETQDQSEVGGAQDQSE 529

```

# RESULT 8

C89824

hypothetical protein sdrC [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C;Accession: C89824

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lian, J.; Ito, T.; Kanamori, M.; Matsumaru, H.; Maruyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hirakawa, H.; Kuhara, S.; Goto, S.; Yabuzaki, J.; Kanehisa, M.; Yamashita, A.; Oshima, K.; Furuya, K.; Yoshino, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: C89824

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-953 <KUR>

A;Cross-references: UNIPROT:Q99W48; UNIPARC:UPI00000CAB80; GB:BA000018;

PID:g13700453; PIDN:BAB41750.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: sdrC

Query Match 8.5%; Score 234; DB 2; Length 953;

Best Local Similarity 23.0%; Pred. No. 7.4e-05;

Matches 131; Conservative 81; Mismatches 216; Indels 142; Gaps 28;

```

Qy      35 YCNSLPDIPFDP---KFITYPF-DQNRVQYKATSLEK----QHKHDLLETPD-LGVTID 85
      : :| | :| | | | | | | | : | : | | :|
Db      375 FVTNLTGYKFNPDAKNFKIYEVTQDQNFVDSFTPDTSKLKDVGTGQFDVIYSNDNKTATVD 434

Qy      86 LINPDITYRIDPNVLLDPADEKLLEEEIQAPTSSKRSQQHAKVVPWMRKTEY-ISTEFNRY 144
      |:| : :| : : : : | :| | : | : :| :| :
Db      435 LLNGQS-----SSDKQYIIQQVAYPDNS--STDNGKI-----DYTLETQNGKS 475

Qy      145 GISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIEKTFEDAQKSISQHYSKPRVTPVEV 204
      || | | | : : | | : || | | : : | |
Db      476 SWSNSYSNVN-GSSTAN--GDQKKYNLGD-----YVWEDTNKDGKQDANEKGIGVYV 525

Qy      205 MPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEMMSQAMIRGMMDEEGN-QF----- 257
      : : : : | | : : | | | |
Db      526 I-----LKDSNG-----KELDRTTTDENGKYQFTGLSNG 554

Qy      258 ---VAYFLPVEETLKKRKRQEEEMDY-----APDDVYD---YKIAR-----EYN 296
      | : | | : : | | : | : :|

```

Db 555 TYSVEFSTPAGYTPPTANAGTDDAVSDSGLTTTGVIKDADNMTLDSGFYKTPKYSLG DYV 614

Qy 297 WNVKNK-----ASKGY-----EENYFFIFREGDGVYVNELETRV 330  
 | | | | | : | | | | : |

Db 615 WYDSNKDGKQDSTEKGIGVKVTLQNEKGEVIGTTETDENGKYRFDNLD SGKY-----KV 669

Qy 331 RLSKRRAKAGV-QSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEE---EEEEMET 386  
 | | | | | : | | | | : | | | | : : : :

Db 670 IFEK---PAGLTQTGTNTTEDDKDADGGEVDVTITDHDDFTLDNGYEEETSDSDSDSDS 726

Qy 387 EEKEAGGSDEEQEKGSSEKEGSEDEHSGSESEERE-EGDRDEASDK---SGSGEDESSD 442  
 : | : : | : : | | : : : | | | | | : :

Db 727 SDS 786

Qy 443 EARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDNDSGSGNGGQRRSHRSRASP 502  
 : : : | : | : | : | : | : | : | : | : | : | : | : |

Db 787 SDS 845

Qy 503 FPSGSEHSAQEDGSEAAASDS-SEADSDSD 531  
 | | : : | : | | | : | | | |

Db 846 SDS 875

# RESULT 9

D89824

hypothetical protein sdrD [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C;Accession: D89824

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lian, J.; Ito, T.; Kanamori, M.; Matsumaru, H.; Maruyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hirakawa, H.; Kuhara, S.; Goto, S.; Yabuzaki, J.; Kanehisa, M.; Yamashita, A.; Oshima, K.; Furuya, K.; Yoshino, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: D89824

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1385 <KUR>

A;Cross-references: UNIPROT:Q99W47; UNIPARC:UPI00000CAA1F; GB:BA000018;

PID:g13700454; PIDN:BAB41751.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: sdrD

Query Match 8.3%; Score 230.5; DB 2; Length 1385;

Best Local Similarity 20.4%; Pred. No. 0.00017;

Matches 116; Conservative 89; Mismatches 236; Indels 127; Gaps 20;

Qy 27 SGVVCrvKYCNslPDIPDPKFITYPFdQNRfVQYKATSLEKQHKHDLlTEPDlGVTIDl 86  
 : | | : | : | : | : | : | : | : |

Db 772 TGVl-----NGADNMTLDSGF--YKTPKYNlGNYVWEDTNKDGKQDSTEKGISGVTVTL 823

Qy 87 INPD-----TYRID-----PNVLLDPADEKLLLEEEIQAP 115



A;Gene: *clfB*

Matches 125; Conservative 76; Mismatches 207; Indels 146; Gaps 23;

Db 271 VDYSNSNNTMPIADIK-----STNGDVVAKAT-----YDILTKTYTFVFTDYVNNKE 317

376 S<sup>1</sup>QIIGVDTASGONTYKOTVF-----VNP<sup>2</sup>KORVLGNTWVYIKGYODKI-E<sup>3</sup>ESSGKVSATDT 429

430 KLRIFEVNDTSKLSDSYYADPNDSNLKEVTDOFKNRIYYEHPNVASIKFGD----- 480

DB 481 --ITKTYVVLVEGHYDNTG-----KNLKQTQVIQENVDPVTNRDYSI---F 520

DB 521 GWNNEN-----VVRYG-----GGSADGDSA----- 540

D6 541 VNPKDPTPGPPVDPEPSPDPEPEPTPDPEPSPDPEPEPSPDPDPDSDSDSDSGSDSDSGS 600

D6 601 DSPSDSPDSPDSPDSPDSPDSESDSDSESDSDSDSDSDSDSDSDSEDSDSDSDSDSDS 660

```
D6      661 DSPSDSPDSPSESDPSDSDSDSDSDSDSD-SDSDSDSDSDSDSDSDSEDSDSDSD 719
```

```

: ||| |::|||||
Db 720 SDSDSDSDSDSDSDSD 733

```

R;McDevitt, D.; Francois, P.; Vaudaux, P.; Foster, T.J.

NID:q397525; PIDN:CAA79304.1; PID:q397526

Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;

DB 780 D 780

A;Note: host *Saimiri sciureus* (common squirrel monkey)

C;Date: 16-Oct-1992 #sequence\_revision 16-Oct-1992 #text\_change 08-Oct-1999  
 C;Accession: A36811  
 R;Albrecht, J.  
 submitted to the EMBL Data Library, January 1992  
 A;Description: Primary structure of the herpesvirus saimiri genome.  
 A;Reference number: A36806  
 A;Accession: A36811  
 A;Molecule type: DNA  
 A;Residues: 1-797 <ALB>  
 A;Cross-references: UNIPARC:UPI00001385A1; GB:X64346; NID:g60320;  
 PIDN:CAA45671.1; PID:g60369  
 R;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.;  
 Newman, C.; Wittmann, S.; Craxton, M.A.; Coleman, H.; Fleckenstein, B.; Honess,  
 R.W.  
 J. Virol. 66, 5047-5058, 1992  
 A;Title: Primary structure of the herpesvirus saimiri genome.  
 A;Reference number: A37309; MUID:92333688; PMID:1321287  
 A;Contents: annotation; protein-coding frames  
 A;Note: neither protein nor nucleotide sequence is given  
 C;Genetics:  
 A;Gene: 48

Query Match 8.1%; Score 224; DB 2; Length 797;  
 Best Local Similarity 23.3%; Pred. No. 0.0002;  
 Matches 91; Conservative 56; Mismatches 181; Indels 62; Gaps 14;

Qy	157	VSVKQQFTEEEI-----YKDRDSQITAIEKTFEDAQKSISQHYSKPRVTPVEVMPVF	208
		:     :   :   :   :   :   :   :   :	
Db	330	VSEYEDFDEDEVELCISDDEVDSEDGNLCVL----DDESESVNS-VALRQVLTVDKQANE	384
Qy	209	PDFKMWINPCAQVIFSDPAPKDTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETL	268
		::    :   :   :   :   :   :   :   :   :	
Db	385	KEYKKIIDKSD----DRDDRDKD-----EYELENEEYNRDEEEDGED-----EEDE	427
Qy	269	KKRKRQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENYFFIFREGDGVYNELET	328
		:   :   :   :   :   :   :   :   :   :	
Db	428	KDEKEGEDEGDDGEDEGED-----EGEDEGDEGDEGDE----GEDEGEDEDEED	474
Qy	329	RVRLSKRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEE	388
		:   :   :   :   :   :   :   :   :	
Db	475	EGEDEGDEGDEGEDEGDEG-----DEGEDEGDEGDEGKDEGDEGDEGDEGDE	525
Qy	389	KEAG--GSDEEQEKGSSSEKEGSEDEHSGSESERE---EGD--RDEASDKSGSGEDESSE	441
		:       : :                       :	
Db	526	GDEGDEGEDEWEDEGDEGEDEGDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGE-GE	584
Qy	442	DEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDNDSDSGSGNGGGQSRSHRSAS	501
		:   :   :   :   :   :   :   :   :	
Db	585	DEGDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDE	642
Qy	502	PFPSGSEHSAQEDGSEAAASDSSEADSDSD	531
		:     :   :	
Db	643	GEDEGDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDE	672

RESULT 13  
 T28679

fibrinogen-binding protein homolog - Staphylococcus aureus

C;Species: Staphylococcus aureus

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T28679

R;Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.J.

Microbiology 144, 3387-3395, 1998

A;Title: Three new members of the serine-aspartate repeat protein multigene family of Staphylococcus aureus.

A;Reference number: Z20510; MUID:99098700; PMID:9884231

A;Accession: T28679

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1315 <JOS>

A;Cross-references: UNIPROT:O86488; UNIPARC:UPI0000052285; EMBL:AJ005646;

NID:e1318791; PID:e1318792; PIDN:CAA06651.1

C;Genetics:

A;Gene: sdrD

Query Match 8.1%; Score 223.5; DB 2; Length 1315;

Best Local Similarity 21.7%; Pred. No. 0.00038;

Matches 121; Conservative 85; Mismatches 225; Indels 127; Gaps 23;

```
Qy      27 SGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQYKATSLEKQHKHDLLEPDLGVTIDL 86
      :||:      |  ::  |  |  |  :      |  |  |  :  |||: |
Db      772 TGV-----NGADNMTLDSGF--YKTPKYNLGNVWEDTNKDGKQDSTEGISGVTVTL 823

Qy      87 INPD-----TYRID-----PNVLLDPADEKLLEEEIQAP 115
      | :      ||::      | :  || ::  :
Db      824 KNEGEVLQTTKTDKDGKYQFTGLENGTYKVEFETPSGYTPTQVSGTDEG-IDSNGTST 882

Qy     116 TSSKRSQQHAKV-----VPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYK 170
      | : : : :      | :|: :|: |: : : :  ||:|      |
Db      883 TGVIKDKDNDTIDSGFYKPTYNLGDYVWEDTNKNGVQDKDEKGISGVTV-----TLK 934

Qy     171 DRDSQITAIEKTFEDAQKSISQ----HYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSD 226
      | : : : | |: : : :      | ||  |      : :
Db      935 DENDKVLKTVTTDENGKYQFTDLNNGTYKVEFETPSGYTPT-----SVTSGN 981

Qy     227 PAPKDTSGAAALEMMSQAMIRGMMDEEG-----NQFVAYFLPVEETLKKRKRQEE 277
      ||::|      :: |  | : |      :||      :: | |:| |
Db      982 DTEKDSNGLTTTGVIKDA--DNMTLDSGFYKTPKYSLGDYVWY-----DSNKDGKQDSTE 1034

Qy     278 EMDYAPDDVYDYKIAREYNWNVKNK--ASKGYEENYFFIFREGDGVYYNELETRVRLSKR 335
      :      : | |:  | | : : :|| : |  |  | :|  |
Db     1035 K-----GIKDVKVTL---LNEKGEVIGTTKTDENGKYCFDNLDGSKY-----KVIFEK- 1079

Qy     336 RAKAGV-QSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGS 394
      ||: |::|||  |  | | : :      |:|  |||  :      |
Db     1080 --PAGLTQTGTNTTEDDKDADGGEVDVTITDHDFTLDNGYEEETSD-----S 1126

Qy     395 DEEQEKGSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEEIF 454
      | : : | |::: |  | |:|: : | | | | | | | : : : | :
Db     1127 DSDSDSDSDSDRSDSDSDSDSDSDSD-SDSDSDSDSD-SDSDSDRSDSDSDSDSDSDSDSDS 1184

Qy     455 GSDADSEDDADSDDEDRGQAQGGSDNDSDSGSGNGGGQSRSHRSASPFPSGSEHSAQED 514
      ||::||: |::|| :      :  ||:|||| |:  |  |  |  | : : |
```

Qv : 337 AKAGVQSGTNALLVVKHRDMNEKELEAQ-EARKAQLNHEPEEEEEEE---EMETEEKEAG 392

```

      : ||:      : | ::|: | : | : | : | ||:|:| | : ||:
Db      529 KD---EEGTD---YEEDTDDSDKDEETKVEEKKTERDEEETEEDKEKETEVEKKKTEKDEE 582
Qy      393 GSDEEQEKGSSSE-----KEGSEDEHSGSE-----SEREEGDRDEASD 430
      |:| |:| : | : :||:| | : |:|: |: |
Db      583 GTDYEEDTDDSDKDVETEVEETDAEDKEENEEGTDDEEDKVEETDLDDQEEEDGEEDKEDD 642
Qy      431 KSGSGEDESSEDEARAARDKEEIFGSD---ADSEDDADSDDEDRGQAQGGSDNDSDSGSN 487
      | ||: |:| : | :| : | ||| : | || : : : | :
Db      643 KEKDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKEDDKED 702
Qy      488 GGGQSRSHRSRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531
      : : :|| | |:| |:|
Db      703 KEKDKEDDKEDK-----KEDNKEKDKEDNKEKDKEDD 734

```

# RESULT 15

A54138

acidic repetitive protein arpl - Tetrahymena thermophila

C;Species: Tetrahymena thermophila

C;Date: 29-Sep-1999 #sequence\_revision 29-Sep-1999 #text\_change 09-Jul-2004

C;Accession: A54138

R;Heinonen, T.Y.; Pearlman, R.E.

J Biol Chem. 269, 17428-17433, 1994

A;Title: A germ line-specific sequence element in an intron in Tetrahymena thermophila.

A;Reference number: A54138; MUID:94292495; PMID:8021245

A;Accession: A54138

A;Molecule type: DNA

A;Residues: 1-334 <HEI>

A;Cross-references: UNIPROT:O77406; UNIPARC:UPI000007FF9E; GB:X76125;

NID:g426479; PIDN:CAA53731.1; PID:e1326004; PID:g3676249

A;Experimental source: strain CU329, macronuclei

A;Note: sequence extracted from NCBI backbone (NCBIN:149332, NCBIP:149333)

C;Genetics:

A;Gene: TAP1

A;Genetic code: SGC5

A;Introns: 64/1; 158/1

Query Match 8.0%; Score 220; DB 2; Length 334;

Best Local Similarity 27.6%; Pred. No. 0.00012;

Matches 64; Conservative 43; Mismatches 85; Indels 40; Gaps 11;

```

Qy      336 RAKAGVQSGTNALLVVKHRDMNE---KELEAQEARKAQLNHEPEE-----EEE 381
      | : :| :| | | :| | | :: | : :| : || :|
Db      24 RKEKPIQKSHSA--VSKETEMTENTPKLIQDDEENADEGDNGDDEESGSDSDSDSGSDDE 81
Qy      382 EEMETEEKEAGGSDEEQEKGSSSEKEG-SEDEHSGSESEREEGDRD----EASDKSG--- 433
      | ::::|:| ||::: | |:| |:| || | : | || | : : | :|
Db      82 ESGSDDEESGSDSDQESGSDDEESGSDDEESGSDDEESGSDSDDDNGSDSDDDNGSDSD 141
Qy      434 --SGEDESSE-----DEARAARDKEEIFGSDADSEDDA-DSDD-EDRGQAQGGSD- 479
      |:|:|:|:| :| | | : | ||: ||| | || | | : | |
Db      142 EDNGDDDSNDDNGDDENGDDAEDGDDAED--GDDAEDGDDAEDGDDAEDGDDAEDGDDA 199
Qy      480 NDSDSGSNGGGQSRSHRSRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531
      | | :| : : | :: ||| :| | :| |::

```

Db 200 EDGDDAEDGDAAEDGDDAEDGDDAEDGDDNEDAEDGDDAEDGDDAEDGDDNE 251

Search completed: April 25, 2006, 09:11:53  
Job time : 45 secs

GenCore version 5.1.7  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2006, 09:03:34 ; Search time 189 Seconds  
 (without alignments)  
 1234.445 Million cell updates/sec

Title: US-10-721-553-2  
 Perfect score: 2764  
 Sequence: 1 MAPTIQTQAQREDGHRPNH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
 1: geneseqp1980s:\*  
 2: geneseqp1990s:\*  
 3: geneseqp2000s:\*  
 4: geneseqp2001s:\*  
 5: geneseqp2002s:\*  
 6: geneseqp2003as:\*  
 7: geneseqp2003bs:\*  
 8: geneseqp2004s:\*  
 9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2764	100.0	531	2	AA42226	Aay42226 Human pan
2	2764	100.0	531	7	ADD18712	Add18712 Human dis
3	2764	100.0	531	8	ADO58688	Ado58688 Human reg
4	2764	100.0	531	8	ABM82102	Abm82102 Tumour-as
5	2744	99.3	531	4	AAB93517	Aab93517 Human pro
6	2658.5	96.2	553	4	ABG19682	Abg19682 Novel hum
7	2464	89.1	473	3	AAB56316	Aab56316 Human sec
8	1244.5	45.0	538	4	ABB59163	Abb59163 Drosophil

9	622	22.5	133	4	ABG19681	Abg19681	Novel hum
10	595	21.5	115	3	AAG03326	Aag03326	Human sec
11	283	10.2	475	4	ABG19412	Abg19412	Novel hum
12	283	10.2	475	8	ADS12265	Ads12265	Human the
13	253	9.2	445	6	ABR53245	Abr53245	Protein s
14	253	9.2	445	7	ADK63670	Adk63670	Disease t
15	237.5	8.6	1633	6	ABU42513	Abu42513	Protein e
16	234.5	8.5	1802	3	AAy83170	Aay83170	Cell wall
17	234.5	8.5	1802	3	AAy70119	Aay70119	Staph. ep
18	234	8.5	953	6	ABU16533	Abu16533	Protein e
19	233.5	8.4	930	2	AAy08641	Aay08641	S. aureus
20	233.5	8.4	947	6	ABJ18940	Abj18940	Pathogen
21	233	8.4	932	4	AAU36845	Aau36845	Staphyloc
22	233	8.4	932	4	AAU34082	Aau34082	Staphyloc
23	232.5	8.4	995	6	ABM72437	Abm72437	Staphyloc
24	230.5	8.3	1385	6	ABU16400	Abu16400	Protein e
25	230	8.3	839	8	ADU02517	Adu02517	Novel hum
26	230	8.3	877	6	ABU42504	Abu42504	Protein e
27	229	8.3	1920	6	ABU43489	Abu43489	Protein e
28	228	8.2	428	5	ABG93245	Abg93245	C. albica
29	227	8.2	567	4	AAE13147	Aae13147	Human ret
30	226	8.2	743	6	ADA89690	Ada89690	Staphyloc
31	226	8.2	877	6	ADA89539	Ada89539	Staphyloc
32	226	8.2	877	6	ABM72702	Abm72702	Staphyloc
33	226	8.2	913	6	ABJ18917	Abj18917	Pathogen
34	225.5	8.2	927	6	ABM72221	Abm72221	Staphyloc
35	225.5	8.2	933	3	AAy58435	Aay58435	Staphyloc
36	225.5	8.2	933	4	AAB69508	Aab69508	Staphyloc
37	225.5	8.2	933	6	ABJ18947	Abj18947	Pathogen
38	225.5	8.2	936	2	AAW89801	Aaw89801	Staphyloc
39	224.5	8.1	194	4	ABG11265	Abg11265	Novel hum
40	224	8.1	265	5	ABG32640	Abg32640	Staphyloc
41	223.5	8.1	1315	2	AAy08642	Aay08642	S. aureus
42	223.5	8.1	1315	6	ABJ18969	Abj18969	Pathogen
43	222	8.0	918	2	AAy08640	Aay08640	S. aureus
44	221.5	8.0	565	9	ADW23812	Adw23812	Staphyloc
45	221.5	8.0	1132	2	AAR97866	Aar97866	Chicken l

#### ALIGNMENTS

##### RESULT 1

AAy42226

ID AAY42226 standard; protein; 531 AA.

XX

AC AAY42226;

XX

DT 20-DEC-1999 (first entry)

XX

DE Human pancreatic differentiation 2 protein sequence.

XX

KW Human; PD2; cancer; regulation; differentiation; neoplastic; therapy;  
KW pancreatic differentiation 2; diagnosis; pancreatic adenocarcinoma;  
KW phosphoprotein.

XX

OS Homo sapiens.

```

XX      WO9950408-A1.
XX
XX      07-OCT-1999.
XX
XX      26-MAR-1999;    99WO-US006633.
XX
XX      27-MAR-1998;    98US-0079649P.
XX
XX      (UYNE-) UNIV NEBRASKA.
XX
XX      Batra SK,  Hollingsworth MA;
XX
XX      WPI; 1999-591317/50.
XX      N-PSDB; AAZ25433.
XX
XX      New phosphoprotein useful as targets for therapy of pancreatic
XX      adenocarcinomas.
XX
XX      Claim 7; Fig 2; 97pp; English.
XX
XX      The present sequence is the human pancreatic differentiation 2 (PD2)
XX      protein, which comprises an amino terminal helix-loop-helix domain and a
XX      centrally localised nuclear transporter signal and nucleotide binding
XX      site. The PD2 nucleotide sequence and a transformed host cell are useful
XX      for screening a test compounds for PD2 modulating activity indicated by
XX      an alteration in the phosphorylation of status of PD2. The host cells are
XX      assessed for altered expression of pancreatic differentiation markers
XX      (MUC-1 or carbonic anhydrase), and modulating activity is correlated with
XX      an alteration in cellular morphology. The PD gene and protein represent
XX      valuable targets in the differential diagnosis and therapy of pancreatic
XX      adenocarcinomas
XX
XX      Sequence 531 AA;
XX

```

Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQEEEMDYAPDDVYDYKIAREYNWNVK 300  
 Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360  
 Qy 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420  
 Qy 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480  
 Qy 481 DSDSGSNGGGQRSRSHRSASPFPSPGSEHSAQEDGSEAAASDSSEADSDSD 531  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 481 DSDSGSNGGGQRSRSHRSASPFPSPGSEHSAQEDGSEAAASDSSEADSDSD 531

RESULT 2

ADD18712

ID ADD18712 standard; protein; 531 AA.

XX

AC ADD18712;

XX

DT 15-JAN-2004 (first entry)

XX

DE Human disease related protein SeqID143.

XX

KW human; disease state; cytostatic; antiinflammatory; ophthalmological;

KW antiarteriosclerotic; vulnerary; gene therapy;

KW hypoxia-regulated condition; tumourigenesis; angiogenesis; apoptosis;

KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;

KW glucose transportation; catecholamine synthesis; iron transport;

KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;

KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;

KW inflammatory condition; wound healing.

XX

OS Homo sapiens.

XX

PN WO2003018621-A2.

XX

PD 06-MAR-2003.

XX

PF 23-AUG-2002; 2002WO-GB003892.

XX

PR 23-AUG-2001; 2001GB-00020558.

PR 05-OCT-2001; 2001GB-00024037.

XX

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX

PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX

DR WPI; 2003-290046/28.

DR N-PSDB; ADD18713.

XX

PT New substantially purified polypeptide, useful for diagnosing or treating



Db 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480

Qy 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531

RESULT 3

ADO58688

ID ADO58688 standard; protein; 531 AA.

XX

AC ADO58688;

XX

DT 15-JUL-2004 (first entry)

XX

DE Human regulatory molecule HRM-9.

XX

KW cytostatic; immunomodulator; agonist; antagonist; gene therapy;

KW human regulatory molecule; HRM; disease development; cell proliferation;

KW immune response; cancer.

XX

OS Homo sapiens.

XX

PN US2002058264-A1.

XX

PD 16-MAY-2002.

XX

PF 26-SEP-2001; 2001US-00840787.

XX

PR 23-SEP-1997; 97US-00933750.

PR 20-JAN-1999; 99US-00234613.

PR 03-MAR-2000; 2000US-00518865.

XX

PA (INCY-) INCYTE PHARM INC.

XX

PI Lal P, Hillman JL, Bandman O, Shah P, Au-Young J, Yue H;

PI Guegler KJ, Corley NC;

XX

DR WPI; 2004-459763/43.

DR N-PSDB; ADO58737.

XX

PT New human regulatory molecules, useful in the diagnosis and treatment of  
 PT cancer and immune disorders.

XX

PS Claim 1; SEQ ID NO 9; 116pp; English.

XX

CC The invention describes human regulatory molecules (HRM) (I) selected  
 CC from a group comprising the fully defined amino acid sequences of SEQ ID  
 CC NOs: 1-49. Also described are: an isolated polynucleotide (II) comprising  
 CC a nucleic acid sequence encoding (I) or the complement of the  
 CC polynucleotide (SEQ ID NOs:50-98); a composition comprising (II) and a  
 CC reporter molecule; an expression vector containing (II); a host cell  
 CC containing the vector; detecting (M1) expression of a nucleic acid in a  
 CC sample; screening (M2) a plurality of molecules to identify a ligand;  
 CC diagnosing (M3) a disease associated with gene expression in a sample  
 CC containing nucleic acids; a composition comprising (I) and a  
 CC pharmaceutical carrier or a labeling moiety; screening (M4) a plurality



AC ABM82102;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Tumour-associated antigenic target (TAT) polypeptide PRO83014, SEQ:5424.  
XX  
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO2004030615-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 29-SEP-2003; 2003WO-US028547.  
XX  
PR 02-OCT-2002; 2002US-0414971P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Wu TD, Zhang Z, Zhou Y;  
XX  
DR WPI; 2004-347921/32.  
DR N-PSDB; ACN40565.  
XX  
PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
PS Claim 12; SEQ ID NO 5424; 7273pp; English.  
XX  
CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence

CC represents a TAT polypeptide of the invention

XX

SQ Sequence 531 AA;

Query Match 100.0%; Score 2764; DB 8; Length 531;

Best Local Similarity 100.0%; Pred. No. 1.5e-196;

Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNCLPDIPFDPKFITYPFDQNRQVQ 60
        |||
Db      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNCLPDIPFDPKFITYPFDQNRQVQ 60

Qy     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
        |||
Db     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120

Qy    121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180
        |||
Db    121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180

Qy    181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
        |||
Db    181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240

Qy    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300
        |||
Db    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy    301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
        |||
Db    301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360

Qy    361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGSEDEHSGSESER 420
        |||
Db    361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGSEDEHSGSESER 420

Qy    421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
        |||
Db    421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480

Qy    481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAAASDSSEADSDSD 531
        |||
Db    481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAAASDSSEADSDSD 531
```

RESULT 5

AAB93517

ID AAB93517 standard; protein; 531 AA.

XX

AC AAB93517;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:12853.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX  
 PS Claim 8; SEQ ID NO 12853; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 531 AA;

Query Match 99.3%; Score 2744; DB 4; Length 531;  
 Best Local Similarity 99.4%; Pred. No. 4.5e-195;  
 Matches 528; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfVQ	60
Db	1	MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRIVQ	60
Qy	61	YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR	120
Db	61	YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR	120
Qy	121	SQQHAKVVPWmrKTEYISTEFNryGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE	180
Db	121	SQQHAKVVPWmrKTEYISTEFNryGISNEKPGVKIGVSVKQQFTEEEIYKDRDSQITAIE	180
Qy	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	240
Db	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	240
Qy	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRdQEEEMDYAPDDVYDYKIAREYNWNVK	300
Db	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRdQEEEMDYAPDDVYDYKIAREYNWNVK	300
Qy	301	NKASKGYEENYFFIFREGDGVYyNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Db	301	NKASKGYEENYFFIFREGDGVYyNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Qy	361	LEAQEARKAQLenHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	420
Db	361	LEAQEARKAQLenHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	420
Qy	421	EEGDRDEASDKSGSGEDESEDEARAARDKEEIfGSDADSEDDADSDDEDRGQAQGGSDN	480
Db	421	EEGDRDEASDKSGSGEDESEDEARAARDKEEIfGSDADSEDDADSDDEDRGQAQGGSDN	480
Qy	481	DSDSGSNGGGQrSRSHSRsAsPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531
Db	481	DSDSGSNGGGQrSRSHSRsAsPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531

# RESULT 6

ABG19682

ID ABG19682 standard; protein; 553 AA.

XX

AC ABG19682;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #19673.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.



Qy	181	KT FEDAQKS-----ISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAP	229
Db	192	KT FEDAQKSVIEGLGWGEARISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAP	251
Qy	230	KDTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDY	289
Db	252	KDTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDY	311
Qy	290	KIAREYNWNVKNKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALL	349
Db	312	KIAREYNWNVKNKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALL	371
Qy	350	VVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGS	409
Db	372	VVKHRDMNEKELEAQETRKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGS	431
Qy	410	EDEHSGSESEREEDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDE	469
Db	432	EDEHSGSESEREEDRHEASDKSGSGQDDSSDYXARAARDKEEIFGSDADSEDDADSDDE	491
Qy	470	DRGQAQGGSDNDSDSGSGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSD	529
Db	492	DRGQAQGGSDNDSDSGRNGGGQRTSRSHRSASPFPSGSEHSAQENGSEAAASDSSEADSD	551
Qy	530	SD 531	
Db	552	SD 553	

# RESULT 7

AAB56316

ID AAB56316 standard; protein; 473 AA.

XX

AC AAB56316;

XX

DT 13-MAR-2001 (first entry)

XX

DE Human secreted protein sequence encoded by gene 106 SEQ ID NO:410.

XX

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; ophthalmological; gene therapy; pathological condition;  
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; skin aging; food additive; preservative.

XX

OS Homo sapiens.

XX

PN WO200070042-A1.

XX

PD 23-NOV-2000.

XX

PF 11-MAY-2000; 2000WO-US012788.



Qy	181	KTfEDAQKSISQHYskPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	240
Db	181	KTfEDAQKSISQHYskPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	240
Qy	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRdQEEEMDYAPDDVYDYKIAREYNWNVK	300
Db	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRdQEEEMDYAPDDVYDYKIAREYNWNVK	300
Qy	301	NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAGVQSGTNALLVVKHRDMNEKE	360
Db	301	NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAGVQSGTNALLVVKHRDMNEKE	360
Qy	361	LEAQEARKAQLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	420
Db	361	LEAQEARKAQLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSXKEGSEDEHSGSESER	420
Qy	421	EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQ	473
Db	421	EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQ	473

RESULT 8

ABB59163

ID ABB59163 standard; protein; 538 AA.

XX

AC ABB59163;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 4281.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR

11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR

N-PSDB; ABL03266.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

XX

PS Disclosure; SEQ ID NO 4281; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 538 AA;

Query Match 45.0%; Score 1244.5; DB 4; Length 538;

Best Local Similarity 50.0%; Pred. No. 1.1e-83;

Matches 271; Conservative 66; Mismatches 172; Indels 33; Gaps 11;

```
Qy      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRVFQ 60
      | ||| | : | :| :||| | :||| | | : ||| :|||
Db      1 MPPTINNSAVNSAAEK-RPQRQTERKSEICrvKYGNNLPDIPFDLKFQYPFDSHRVFQ 59

Qy     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
      | ||| : | :||| ||| :||| : | | ||| | | | | | | | | | |
Db     60 YNPTSLERNFKYDVLTEHDLGVTVDLINRELYQADSMTLDPADKLEEEETLTPTDSVR 119

Qy    121 SQQHAKVVPWMRKTEYISTEFNRYGISN-EKPEVKIGVSVKQQFTEEIYKDRDSQITAI 179
      | :|| : | | :|| :||| | : | | | | :| :| :| :| | |
Db    120 SRQHSRTVSWLRKSEYISTEQTRFQPQNLENIEAKVGYNVKKSRLREETLYLDREAQIKAI 179

Qy    180 EKTFFEDAQKSISQHYSKPRVTPVEVMPVFDFKMWINPCAQVIFSDPAPKDTSGAAALE 239
      ||| | : | :||| | ||| :| :||| | ||| | | | | | : | |
Db    180 EKTFSDTKSEITKHYSKPNVVPVEVLPIFPDFTNWKFPQVIFSDPAPAGKNVPAQLE 239

Qy    240 MMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYDYKIAREYNWNV 299
      ||| | | :|| | ||| | | :|| :| | | | | :| :| :| :| | |
Db    240 EMSQAMIRGVMDESGEQFVAYFLPTEQTLEKRRTDFINGELYKEEEYKYIAREYNWNV 299

Qy    300 KNKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEK 359
      | ||| | | | | : | :||| | | | :|| | | | | | | | : :
Db    300 KTKASKGYEENYFFVMRQ-DGIYYNELETRVRLNKRVRKVG-QQPNNTKLTVKHRPLDSM 357

Qy    360 ELEAQEARKAQLNHEPEEE-----EEEEM----ETEE-----KEAGGSD----- 395
      | | | : || | | | | | | | : | | : | | : |
Db    358 EHRMQRYRERQLEVPGEIEEVEEVEEQMQIIGETEKTSEDA AVGAQAASGADSPAQV 417

Qy    396 --EEQKGSSEKEGSEDEHSGSESEREEDRDEASDKSGSGEDESSEDEARAARDKEEI 453
      : | : | : || || | | : : : ||| : | : :
Db    418 ARDRQSRSRSTRSGS-SSGSGSGSGSRASSRSKSGSRSGSGSRRTNSPAGSQKSGSR- 475

Qy    454 FGSDADSEDDADSDDEDRGQAQGGSDNDSDSGS-NGGGQSRSHRSASPFPSGSEHSAQ 512
      | : | : | | : : | : | ||| :| | | | ||| | | | :
Db    476 SRSVSRSRSRKSGSRSRSRSRKSGSRSGSRSGSRSPSRSRSGSPSGSGSSSGSA 535

Qy    513 ED 514
      |
```

## RESULT 9

ABG19681

ID ABG19681 standard; protein; 133 AA.

XX

AC ABG19681;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #19672.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS83868.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX

PS Claim 20; SEQ ID NO 50040; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 133 AA;

Query Match 22.5%; Score 622; DB 4; Length 133;  
Best Local Similarity 64.4%; Pred. No. 2.9e-38;  
Matches 130; Conservative 0; Mismatches 2; Indels 70; Gaps 1;

```
QY      273 RDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENYFFIFREGDGVYNELETRVRL 332
          |||||||||||||||||||||||||||||||||||
Db       1  RDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGY----- 35

QY      333 SKRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAG 392
          |||||||||||||||
Db       36 -----EEEEEMETEEKEAG 50

QY      393 GSDEEQEKGSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEE 452
          || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       51 GSYEEQEKGSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEE 110

QY      453 IFGSDADSEDDADSDDEDRGQA 474
          ||||||||||| |||||
Db       111 IFGSDADSEDDADSYDEDRGQA 132
```

RESULT 10

AAG03326

ID AAG03326 standard; protein; 115 AA.

XX

AC AAG03326;

XX

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein, SEQ ID NO: 7407.

XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.

XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.

XX

PF 21-FEB-2000; 2000EP-00200610.

XX

PR 26-FEB-1999; 99US-0122487P.

XX

PA (GEST ) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX

DR WPI; 2000-500381/45.



XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS83599.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 49771; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 475 AA;

Query Match 10.2%; Score 283; DB 4; Length 475;  
Best Local Similarity 87.7%; Pred. No. 2.3e-12;  
Matches 57; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 328 TRVRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLENHEPEEEEEEMETE 387  
:|||||||||||||||||||||||||||||||||||||||||||||||||||||||:  
Db 3 SRVRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLENHEPEEEEEEI RQP 62  
Qy 388 EKEAG 392  
|: |  
Db 63 RKKLG 67

RESULT 12  
ADS12265

ID ADS12265 standard; protein; 475 AA.  
 XX  
 AC ADS12265;  
 XX  
 DT 16-DEC-2004 (first entry)  
 XX  
 DE Human therapeutic contig protein - SEQ ID 2502.  
 XX  
 KW antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnerary;  
 KW inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;  
 KW aplastic anaemia; cancer; wound healing; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1. .475  
 FT /label= Unknown, OTHER  
 FT /note= "OTHER = In-frame STOP codon"  
 XX  
 PN WO2004080148-A2.  
 XX  
 PD 23-SEP-2004.  
 XX  
 PF 30-SEP-2003; 2003WO-US030720.  
 XX  
 PR 02-OCT-2002; 2002US-0416186P.  
 XX  
 PA (NUVE-) NUVELO INC.  
 XX  
 PI Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;  
 PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;  
 XX  
 DR WPI; 2004-668857/65.  
 DR N-PSDB; ADS11667.  
 XX  
 PT New polynucleotide, useful in preparing a composition for diagnosing or  
 PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,  
 PT aplastic anemia or cancer for promoting wound healing.  
 XX  
 PS Example 2; SEQ ID NO 2502; 718pp; English.  
 XX  
 CC The invention relates to a novel isolated polynucleotide and the encoded  
 CC polypeptide. The molecules of the invention demonstrate antiinflammatory,  
 CC neuroprotective, antianaemic, cytostatic and vulnerary activities and may  
 CC be useful in preparing a composition for diagnosing or treating  
 CC inflammatory, haematopoietic, immune, neurodegenerative or stem cell  
 CC disorders, such as aplastic anaemia or cancer, as well as for promoting  
 CC wound healing. The molecules may also be utilised during gene therapy  
 CC procedures. The current sequence is that of a human therapeutic contig  
 CC protein of the invention.  
 XX  
 SQ Sequence 475 AA;  
  
 Query Match 10.2%; Score 283; DB 8; Length 475;  
 Best Local Similarity 87.7%; Pred. No. 2.3e-12;  
 Matches 57; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 328 TRVRLSKRRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLENHEPEEEEEEMETE 387  
 :|||||||||||||||||||||||||||||||||||||||||||||||||||||:  
 Db 3 SRVRLSKRRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLENHEPEEEEEEEIRQP 62

Qy 388 EKEAG 392  
 | : |  
 Db 63 RKKLG 67

RESULT 13

ABR53245

ID ABR53245 standard; protein; 445 AA.

XX

AC ABR53245;

XX

DT 20-JUN-2003 (first entry)

XX

DE Protein sequence #SEQ ID 1355.

XX

KW Multiprotein complex; eukaryote; drug target; diagnosis.

XX

OS Saccharomyces cerevisiae.

XX

PN EP1258494-A1.

XX

PD 20-NOV-2002.

XX

PF 20-DEC-2001; 2001EP-00130253.

XX

PR 15-MAY-2001; 2001EP-00111774.

XX

PA (CELL-) CELLZOME AG.

XX

PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

PI Marzioch M, Schultz JD, Superti-Furga GD;

XX

DR WPI; 2003-250078/25.

DR N-PSDB; ACC61287.

XX

PT New isolated protein complexes useful for diagnosing a disease or  
 PT disorder, or as a target for an active agent of a pharmaceutical,  
 PT preferably a drug target in the treatment or prevention of disease or  
 PT disorder.

XX

PS Disclosure; SEQ ID NO 1355; 17pp + Sequence Listing; English.

XX

CC The invention relates to multiprotein complexes from eukaryotes. Proteins  
 CC of the invention and DNA sequences encoding them are given in records  
 CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are  
 CC obtainable by using a protein as a bait and isolating the set of proteins  
 CC which is attached thereto from cells. Such protein complexes may comprise  
 CC up to 30 distinct proteins. Protein complexes of the invention are useful  
 CC for diagnosing a disease or disorder, or as a target for an active agent  
 CC of a pharmaceutical, preferably a drug target in the treatment or  
 CC prevention of a disease or disorder. Note: The sequence data for this  
 CC patent is not represented in the printed specification, but is based on  
 CC sequence information supplied by the European Patent Office. The complete

CC document is available on CD-ROM

XX

SQ Sequence 445 AA;

Query Match 9.2%; Score 253; DB 6; Length 445;

Best Local Similarity 22.1%; Pred. No. 3.6e-10;

Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;

```
Qy      23 LPERSGVVCRVKYCNSLPDIPFDPKFITYP-----FDQNRVQYKATSLEKQHK 71
      : :: : :|| ||| || : || : :: | : |
Db      1 MSKKQEYIAPIKYQNSLPVPQLPPKLLVYPESPETNADSSQLINSLYIKTNVTNLIQQ-- 58

Qy      72 HDLLTEPDLGVITIDLI-----NPDYRIDPNVLLDPADEKLLEEEIQAPTSSKRS 121
      : ||| : ||| : | | | | | | : | : :
Db      59 -----DEDLGMPVDLMKFPGLLNKLD SKLLYGFD-NVKLDKDDRILLRD-----PRIDRLT 108

Qy     122 QQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQ----- 175
      : | ::|:|:|:| : :: | : :: | ||
Db     109 KTDISKVTFLLRTEYVSNTIAAHNTSLKRKRRL-----DDGSDDENLDV 154

Qy     176 ---ITAIEKTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDT 232
      | : :| || | || | | : : || || |
Db     155 NHIISRVEGTFNKTDK--WQHPVKKGVMVKKWDL LPD----TASMDQVYF-----ILKF 203

Qy     233 SGAAALEMMSQAMIR-GM---MDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYD 288
      |:|:|: : : |: :: | : :: : : : : |: || || ::
Db     204 MGSASLDTKEKSLNTGIFRPVELEDEWISMYATDHKDSAILENELEKGMDEMDDDSHE 263

Qy     289 ---YKIAREYNWNVKNKASKGYEENYFFIFREGDGV-YYNELETRVRLSKRRKAGVQSG 344
      || | ::::| | | : ||: || | :: : | :| : :
Db     264 GKIYKFKRIRDYDMKQVAEKPMTE-LAIRLNDKDGIAYYKPLRSKIELRRRRVNDI IKP- 321

Qy     345 TNALLVVKH-----RDMNEKELEAQEARKAQLN-----HEPEEEEEEMETEEK 389
      || :| : : || : : || : : : | ::|:|:| : |
Db     322 -----LVKEHDIDQLNVTLRNPSTKEANIRDKLRMKFDPINFATVDEEDDEDEEQPEDVKK 377

Qy     390 EAGGSDEEQEKGSSEKEGSEDEHSGSESEREEDRDEASDKSGSGEDESSEDEARAARD 449
      |: | : : :|| | | :|| | | : : ::|:|:| ||
Db     378 ESEG--DSKTEGSEQEGENEKDEEIKQEKENEQ-----DEENKQDENRAADT 422

Qy     450 KEEIFGSDADSEDDADSDDEDRGQAQ 475
      | ||| : :::: | :
Db     423 PET---SDAVHTEQKPEEEKETLQEE 445
```

#### RESULT 14

ADK63670

ID ADK63670 standard; protein; 445 AA.

XX

AC ADK63670;

XX

DT 06-MAY-2004 (first entry)

XX

DE Disease treating protein complex-derived protein #818.

XX

KW protein complex; drug target; diagnosis.

XX

OS Unidentified.  
XX  
PN EP1338608-A2.  
XX  
PD 27-AUG-2003.  
XX  
PF 20-DEC-2002; 2002EP-00102902.  
XX  
PR 20-DEC-2001; 2001EP-00130253.  
XX  
PA (CELL-) CELLZOME AG.  
XX  
PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;  
PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;  
PI Michon A, Leutwein C, Rick J;  
XX  
DR WPI; 2003-638460/61.  
DR N-PSDB; ADK63671.  
XX  
PT New proteins and protein complexes from eukaryotes, useful as targets in  
PT drug screening, or in diagnosing or screening for the presence of a  
PT disease or disorder, or a predisposition for developing a disease or  
PT disorder in a subject.  
XX  
PS Disclosure; SEQ ID NO 1635; 13pp; English.  
XX  
CC The invention relates to novel protein complexes comprising a first and a  
CC second protein, or its derivative, fragment, homologue or variant. The  
CC proteins are selected from given protein complexes, which are not defined  
CC in the specification. The variants are encoded by nucleic acids that  
CC hybridize to the nucleic acids encoding the proteins under low stringency  
CC conditions. The protein complexes are useful as targets for an active  
CC agent of a pharmaceutical. These protein complexes are particularly  
CC useful as drugs targets for the treatment or preventing of a disease or  
CC disorder. The complexes and methods above are useful in diagnosing or  
CC screening for the presence of a disease or disorder or a predisposition  
CC for developing a disease or disorder in a subject. These are also useful  
CC in screening for a drug for treatment or prevention of a disease or  
CC disorder. The molecule that modulates the amount, activity or protein  
CC components of the complex is useful for the manufacture of a medicament  
CC for the treatment or prevention of a disease or disorder. This sequence  
CC corresponds to a protein of the invention. (Note: the sequence data for  
CC this patent did not form part of the printed specification but was  
CC obtained from the EPO in electronic format).  
XX  
SQ Sequence 445 AA;

Query Match 9.2%; Score 253; DB 7; Length 445;  
Best Local Similarity 22.1%; Pred. No. 3.6e-10;  
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;

Qy 23 LPERSGVVCRVKYCNSLPDIPFDPKFITYP-----FDQNRFFVQYKATSLEKQHK 71  
: :: : || ||| || : || : :: |:| :|  
Db 1 MSKKQEYIAPIKYQNSLPVPQLPPKLLVYPESPETNADSSQLINSLYIKTNVTNLIQQ-- 58  
Qy 72 HDLLTEPDLGVTIDLI-----NPDTYRIDPNVLLDPADEKLLLEEIQAPTSSKRS 121  
: |||: :||: : | || || | || : | : :

Db 59 -----DEDLGMPVDLMKFPGLLNKLDKSKLLYGFD-NVKLDKDDRILLRD-----PRIDRLT 108  
 Qy 122 QQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQ----- 175  
 : | ::|:|::| : :: | : :: | ||  
 Db 109 KTDISKVTFLRRTEYVSNTIAAHDNTSLKRKRRL-----DDGSDDENLDV 154  
 Qy 176 ---ITAIEKTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDT 232  
 |: :| || | || | | | : : || || |  
 Db 155 NHIISRVEGTFNKTDK--WQHPVKKGVKMKVKKWDLPLD----TASMDQVYF-----ILKF 203  
 Qy 233 SGAAALEMMSQAMIR-GM---MDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYD 288  
 |:|:|: : : |: :: | :::: : : : : |: || || ::  
 Db 204 MGSASLDTKEKKS LNTGIFRPVELEEDEWISMYATDHKDSAILENELEKGMDEMDDDSHE 263  
 Qy 289 ---YKIAREYNWNVKNKASKGYEENYFFIFREGDGV-YYNELETRVRLSKRRAKAGVQSG 344  
 || | ::::| | | : ||: || | ::: | :|| ::  
 Db 264 GKIYKFKRIRDYDMKQVAEKPMTE-LAIRLNDKDGIAYYKPLRSKIELRRRRVNDIIP- 321  
 Qy 345 TNALLVVKH-----RDMNEKELEAQEARKAQLN-----HEPEEEEEEMETEEK 389  
 || :| | : : || :: : : : | ::|:|: | :|  
 Db 322 ----LVKEHDIDQLNVTLRNPSTKEANIRDKLRMKFDPINFATVDEEDDEDEEQPEDVKK 377  
 Qy 390 EAGGSDEEQEKGSSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARD 449  
 |: | : : :|| | | :|| | | | : ::|: :|| |||  
 Db 378 ESEG--DSKTEGSEQEGENEKDEEIKQEKENEQ-----DEENKQDENRAADT 422  
 Qy 450 KEEIFGSDADSEDDADSDDEDRGQAQ 475  
 | ||| : :::: | :  
 Db 423 PET---SDAVHTEQKPEEEKETLQEE 445

RESULT 15

ABU42513

ID ABU42513 standard; protein; 1633 AA.

XX

AC ABU42513;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #28040.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Staphylococcus epidermidis.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

DR N-PSDB; ACA46383.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 25; SEQ ID NO 70437; 1766pp; English.

XX

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 1633 AA;

Query Match 8.6%; Score 237.5; DB 6; Length 1633;

Best Local Similarity 21.2%; Pred. No. 2.8e-08;

Matches 131; Conservative 85; Mismatches 238; Indels 165; Gaps 25;

Qy 48 FITYFPDQNRQVQYKATSLEKQHKHDLLEPDLGVTIDLINPDYRIDP-----NVLLDP 102

Db 707 YVTLKDSNNRELQRVTTDQSGHYQFDNLQNGT--YTVEFAIPDNYTPSPANNSTNDAIDS 764

Qy 103 ADEKLLEEEIQAPTSSKRSQQHAKV-----VPWMRKTEYISTEFNRYGI--SNEKPEVK 154

	: : : :     : : : :	
Db	765 DGERDGTRKVVVAKGTINNADNMTVDTGfYlTPKYNVGDYVWEDTNKdGIQDDNEKGISN	824
Qy	155 IGVSVKQ-----QFTEEEIY-----KDRDSQ	175
	: :  :      :	
Db	825 VKVTLKNKNGDTIGTTTTDSNGKYEFTGLENGDYTIEFETPEGYTPTKQNSGSDEGKDSN	884
Qy	176 ITAIEKTFEDA-QKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDP-----	227
	:    :  :    :  :  :	
Db	885 GKTkTVTVKDADNKtIDSGfYKPIYN-----LGdY-VWEDTNKdGIQDDSEKGISGVK	936
Qy	228 -APKDTSGAA----ALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRdQEEEMDY-	281
	:    : :  :  :  :  :  :  :	
Db	937 VTLKDKNGNAIGTTTTDASGHYQfKGL--ENGsYtVEfETPSGYtPTKANSgQDITVDSN	994
Qy	282 -----APDDVYD---YKIAR---EYNWNVKNK-----ASKGY-----	307
	:     : :	
Db	995 GITTTGIINGADNLTIDSGfYKtPKYSVGdYVWEDTNKdGIQDDNEKGISGVKvTLKDEK	1054
Qy	308 -----EENYFFIFREGD-GVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMN	357
	:    :         :  :  : :	
Db	1055 GNIISTTTTtDENGKYQfDNLDsGNYIIHFekPEGmtQTtANSg-----NDD	1100
Qy	358 EKELEAQEARKAQLENH-----EPEEEEEEMETEEKEAGGSd	395
	: : :  : :  : : : : :	
Db	1101 EKdADGEDVR-VtITdHdDfSIDNGYfDDdSDSDSDADSDSDSDSDADSDSDADSNsD	1159
Qy	396 EEQEKGSsSEKEGSedeHSGSESERE-EGDRDEASDK-SGSgeDEsEDEARAAARDKEEI	453
	: :     : :     :  : :           : : :   :	
Db	1160 SDSDSDSDSDSDSDSDSDSDSDSDSDADSDSDADSDSDSDSDSDSDSDSDSDSDSDSDSD	1219
Qy	454 FGSDADsEDDADSDDEDRGQAQGSdNDSDSGSNGGgQRSRSHSRsASPfPSGSEHSAQE	513
	:      : :   :       :             : :	
Db	1220 SDSDADSDSDADSDSDADSDSDADSDSDSDSDSDAD---SDSDSDSDSDADSDSDSDSDS	1276
Qy	514 DGSEAAASDs-SEADSDSD	531
	:       :	
Db	1277 DADSDSDSDSDSDADSDSD	1295

Search completed: April 25, 2006, 09:07:10  
Job time : 194 secs

GenCore version 5.1.7  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2006, 09:04:03 ; Search time 231 Seconds  
 (without alignments)  
 1621.799 Million cell updates/sec

Title: US-10-721-553-2  
 Perfect score: 2764  
 Sequence: 1 MAPTIQTQAQREDGHRPNSH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : UniProt\_05.80:\*  
 1: uniprot\_sprot:\*  
 2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2764	100.0	531	2 Q9H166_HUMAN	Q9h166 homo sapien
2	2744	99.3	531	2 Q9NUU9_HUMAN	Q9nuu9 homo sapien
3	2740	99.1	533	2 Q5RAX0_PONPY	Q5rax0 pongo pygma
4	2730.5	98.8	534	2 Q5RE77_PONPY	Q5re77 pongo pygma
5	2718	98.3	535	2 Q8K2T8_MOUSE	Q8k2t8 mus musculu
6	2708	98.0	535	2 Q9JJ99_MOUSE	Q9jj99 mus musculu
7	2705	97.9	535	2 Q4V886_RAT	Q4v886 rattus norv
8	2531.5	91.6	510	2 O75239_HUMAN	O75239 homo sapien
9	2183.5	79.0	520	2 Q6P2Y1_XENTR	Q6p2yl xenopus tro
10	2030	73.4	503	2 Q4U0S5_BRARE	Q4u0s5 brachydanio
11	1995	72.2	485	2 Q8N7H5_HUMAN	Q8n7h5 homo sapien
12	1984	71.8	377	2 Q9CS63_MOUSE	Q9cs63 mus musculu
13	1935	70.0	407	2 Q68F51_XENLA	Q68f51 xenopus lae
14	1757.5	63.6	370	2 Q4RRR2_TETNG	Q4rrr2 tetraodon n
15	1244.5	45.0	538	2 Q9VN55_DROME	Q9vn55 drosophila

16	1129	40.8	468	2	Q7PXA3_ANOGA	Q7pxa3	anopheles g
17	713	25.8	453	2	Q60MA7_CAEBR	Q60ma7	caenorhabdi
18	645	23.3	425	2	P90783_CAEEL	P90783	caenorhabdi
19	481.5	17.4	499	2	Q55E33_DICDI	Q55e33	dictyosteli
20	372.5	13.5	589	2	Q8RW91_ARATH	Q8rw91	arabidopsis
21	361.5	13.1	593	2	Q9MA04_ARATH	Q9ma04	arabidopsis
22	358.5	13.0	451	2	Q6ZD92_ORYSA	Q6zd92	oryza sativ
23	335.5	12.1	547	2	Q9CA82_ARATH	Q9ca82	arabidopsis
24	311	11.3	703	2	Q4P5K3_USTMA	Q4p5k3	ustilago ma
25	286.5	10.4	386	2	Q6C509_YARLI	Q6c509	yarrowia li
26	280.5	10.1	457	2	Q9US06_SCHPO	Q9us06	schizosacch
27	266.5	9.6	572	2	Q56PB7_RAT	Q56pb7	rattus norv
28	265.5	9.6	791	2	Q9DGL1_FUGRU	Q9dgl1	fugu rubrip
29	263.5	9.5	467	2	Q59Y36_CANAL	Q59y36	candida alb
30	257.5	9.3	571	2	Q8MTN7_TRISP	Q8mtn7	trichinella
31	253	9.2	445	1	PAF1_YEAST	P38351	saccharomyc
32	251.5	9.1	538	2	Q9ET15_MOUSE	Q9et15	mus musculu
33	250.5	9.1	458	2	Q6BT93_DEBHA	Q6bt93	debaryomyce
34	249	9.0	440	2	Q55S74_CRYNE	Q55s74	cryptococcu
35	249	9.0	440	2	Q5KGM6_CRYNE	Q5kgm6	cryptococcu
36	245.5	8.9	784	2	Q7LZ90_TORCA	Q7lz90	torpedo cal
37	245	8.9	1848	2	Q7RGP8_PLAYO	Q7rgp8	plasmodium
38	244.5	8.8	934	2	Q9GMD3_BOVIN	Q9gmd3	bos taurus
39	243.5	8.8	438	2	Q6FXJ9_CANGA	Q6fxj9	candida gla
40	241	8.7	792	2	Q9YTL7_9HERP	Q9yt17	ateline her
41	240	8.7	1394	1	CNGB1_BOVIN	Q28181	bos taurus
42	239.5	8.7	544	2	Q56PC0_FELCA	Q56pc0	felis silve
43	238.5	8.6	1381	2	Q5HIB3_STAAC	Q5hib3	staphylococ
44	238	8.6	613	2	Q6UDM5_BRARE	Q6udm5	brachydanio
45	236.5	8.6	1633	2	Q8CMP4_STAEP	Q8cmp4	staphylococ

# ALIGNMENTS

## RESULT 1

### Q9H166\_HUMAN

ID Q9H166\_HUMAN PRELIMINARY; PRT; 531 AA.  
AC Q9H166;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Hypothetical protein PD2.  
GN Name=PD2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Batra S.K., Choudhury A., Keita M., Schmied B.M., Vanlith M.,  
RA Walter N.A.R., Jokerst J., Sikela J.M., Metzgar R.S.,  
RA Hollingsworth M.A.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Muscle, and Placenta;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Muscle;  
 RG NIH MGC Project;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ401156; CAC20564.1; -; mRNA.  
 DR EMBL; BC013402; AAH13402.1; -; mRNA.  
 DR EMBL; BC000017; AAH00017.1; -; mRNA.  
 DR Ensembl; ENSG00000006712; Homo sapiens.  
 DR InterPro; IPR007133; Paf1.  
 DR Pfam; PF03985; Paf1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 531 AA; 59976 MW; 756F800AA64255D6 CRC64;

Query Match 100.0%; Score 2764; DB 2; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-122;  
 Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAPT	IQTQAQ	REDGHR	PNSHRT	LPERSG	VVCRV	KYCNSL	PDIPFD	PKFIT	YPFDQ	NRFVQ	60
Db	1	MAPT	IQTQAQ	REDGHR	PNSHRT	LPERSG	VVCRV	KYCNSL	PDIPFD	PKFIT	YPFDQ	NRFVQ	60
Qy	61	YKAT	SLEKQ	HKHDL	LTEPDL	GV	TIDLIN	PD	TYRID	PNVLL	DP	AD	120
Db	61	YKAT	SLEKQ	HKHDL	LTEPDL	GV	TIDLIN	PD	TYRID	PNVLL	DP	AD	120
Qy	121	SQ	HAKV	VP	W	MRKTE	YISTE	FN	RGIS	NEKPE	VKIG	SV	180
Db	121	SQ	HAKV	VP	W	MRKTE	YISTE	FN	RGIS	NEKPE	VKIG	SV	180

Qy 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240  
 |||  
 Db 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240

Qy 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300  
 |||  
 Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360  
 |||  
 Db 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360

Qy 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420  
 |||  
 Db 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420

Qy 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480  
 |||  
 Db 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480

Qy 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
 |||  
 Db 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531

## RESULT 2

### Q9NUU9\_HUMAN

ID Q9NUU9\_HUMAN PRELIMINARY; PRT; 531 AA.  
 AC Q9NUU9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein FLJ11123.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Placenta;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,  
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,  
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
 RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,

RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 DR EMBL; AK001985; BAA92020.1; -; mRNA.  
 DR InterPro; IPR007133; Paf1.  
 DR Pfam; PF03985; Paf1; 1.  
 SQ SEQUENCE 531 AA; 59797 MW; 446D2588B20E42DC CRC64;

Query Match 99.3%; Score 2744; DB 2; Length 531;  
 Best Local Similarity 99.4%; Pred. No. 2.3e-121;  
 Matches 528; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MAPT	IQTQAQ	REDGHR	PNSHRT	LPERSG	VVCRV	KYCNSL	PDIPFD	PKFITY	PF	QNR	FVQ	60
Db	1	MAPT	IQTQAQ	REDGHR	PNSHRT	LPERSG	VVCRV	KYCNSL	PDIPFD	PKFITY	PF	QNR	RIVQ	60
Qy	61	YKAT	SLEKQ	HKHDL	LTEPDL	GV	TIDLIN	PD	TYRIDP	NVLLD	PADEK	LLEEEI	QAPTSS	KR 120
Db	61	YKAT	SLEKQ	HKHDL	LTEPDL	GV	TIDLIN	PD	TYRIDP	NVLLD	PADEK	LLEEEI	QAPTSS	KR 120
Qy	121	SQQH	AKVVP	WMRKTE	YISTEF	NRYGIS	NEKPEV	KIGVSV	KQQFTE	EEIYK	DRDSQ	ITAIE	180	
Db	121	SQQH	AKVVP	WMRKTE	YISTEF	NRYGIS	NEKPGV	KIGVSV	KQQFTE	EEIYK	DRDSQ	ITAIE	180	
Qy	181	KTFE	DAQKS	ISQHY	SKPRVT	PVEVMP	VPDFK	MWINP	CAQVI	FSDP	PAPKDT	SGAAALEM	240	
Db	181	KTFE	DAQKS	ISQHY	SKPRVT	PVEVMP	VPDFK	MWINP	CAQVI	FSDP	PAPKDT	SGAAALEM	240	
Qy	241	MSQA	MIRGMM	DEEGNQ	FVAYFL	PVEETL	KKRKR	RDQEE	EMDYAP	DDVYDY	KIAREY	NWNVK	300	
Db	241	MSQA	MIRGMM	DEEGNQ	FVAYFL	PVEETL	KKRKR	RDQEE	EMDYAP	DDVYDY	KIAREY	NWNVK	300	
Qy	301	NKAS	KGYEEN	YFFIF	FREGD	GVYYNE	LETRV	RLSKRR	AKAGV	QSGTN	ALLVVK	HRDMNEKE	360	
Db	301	NKAS	KGYEEN	YFFIF	FREGD	GVYYNE	LETRV	RLSKRR	AKAGV	QSGTN	ALLVVK	HRDMNEKE	360	
Qy	361	LEAQ	EARKA	QLENH	EP	EEEEEE	EMETEE	KEAGGS	DEEQE	KGSSE	KEGSE	DEHSGSE	SER 420	
Db	361	LEAQ	EARKA	QLENH	EP	EEEEEE	EMETEE	KEAGGS	DEEQE	KGSSE	KEGSE	DEHSGSE	SER 420	
Qy	421	EEGD	RDEAS	DKSGS	GEDESSE	DEARA	ARDKEE	IFGSD	ADSEDD	ADSDDE	DRGQAQ	GGSDN	480	
Db	421	EEGD	RDEAS	DKSGS	GEDESSE	DEARA	ARDKEE	IFGSD	ADSEDD	ADSDDE	DRGQAQ	GGSDN	480	

Qy 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
|||||  
Db 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531

RESULT 3

Q5RAX0\_PONPY

ID Q5RAX0\_PONPY PRELIMINARY; PRT; 533 AA.  
AC Q5RAX0;  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE Hypothetical protein DKFZp469K121.  
GN Name=DKFZp469K121;  
OS Pongo pygmaeus (Orangutan).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Pongo.  
OX NCBI\_TaxID=9600;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RG The German cDNA Consortium;  
RA Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,  
RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;  
RL Submitted (NOV-2004) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; CR858891; CAH91090.1; -; mRNA.  
DR InterPro; IPR007133; Paf1.  
DR Pfam; PF03985; Paf1; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 533 AA; 60248 MW; 5D01B4E99420F050 CRC64;

Query Match 99.1%; Score 2740; DB 2; Length 533;  
Best Local Similarity 99.2%; Pred. No. 3.6e-121;  
Matches 529; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQ 60  
|||||  
Db 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQ 60  
  
Qy 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDITYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120  
|||||  
Db 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDITYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120  
  
Qy 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGSVKQQFTEEEIYKDRDSQITAIE 180  
|||||  
Db 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGSVKQQFTEEEIYKDRDSQITAIE 180  
  
Qy 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240  
|||||  
Db 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240  
  
Qy 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQEEEMDYAPDDVYDYKIAREYNWNVK 300  
|||||  
Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQEEEMDYAPDDVYDYKIAREYNWNVK 300  
  
Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360

```

Db      301  |||||
Qy      361  LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEKEKGSSEKEGSEDEHSGSESER 420
Db      361  |||||
Qy      421  EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
Db      421  |||||
Qy      481  DSDSGSNGGGQSRSH--SRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531
Db      481  |||||

```

#### RESULT 4

##### Q5RE77\_PONPY

```

ID  Q5RE77_PONPY PRELIMINARY;      PRT;    534 AA.
AC  Q5RE77;
DT  01-FEB-2005 (TrEMBLrel. 29, Created)
DT  01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT  01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE  Hypothetical protein DKFZp468J227.
GN  Name=DKFZp468J227;
OS  Pongo pygmaeus (Orangutan).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC  Pongo.
OX  NCBI_TaxID=9600;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RC  TISSUE=Heart;
RG  The German cDNA Consortium;
RA  Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA  Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL  Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
DR  EMBL; CR857657; CAH89930.1; -; -.
DR  InterPro; IPR007133; Paf1.
DR  Pfam; PF03985; Paf1; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 534 AA; 60364 MW; E256C4EE1CA171FB CRC64;

```

```

Query Match          98.8%; Score 2730.5; DB 2; Length 534;
Best Local Similarity 99.1%; Pred. No. 1e-120;
Matches 529; Conservative 0; Mismatches 2; Indels 3; Gaps 2;

```

```

Qy      1  MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfVQ 60
Db      1  MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfVQ 60
Qy     61  YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
Db     61  YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
Qy    121  SQQHAKVVPWMrKTEYISTEFNryGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180

```

Db 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180

Qy 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240  
 |||

Db 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240

Qy 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300  
 |||

Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy 301 NKASKGYEENYFFIFREGDGVYNELET-RVRLSKRRKAGVQSGTNALLVVKHRDMNEK 359  
 |||

Db 301 NKASKGYEENYFFIFREGDGVYNELETRRVRLSKRRKAGVQSGTNALLVVKHRDMNEK 360

Qy 360 ELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESE 419  
 |||

Db 361 ELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESE 420

Qy 420 REEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSD 479  
 |||

Db 421 REEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSD 480

Qy 480 NDSDSGSNGGGQRSRSH--SRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
 |||

Db 481 NDSDSGSNGGGQRSRSHSRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 534

# RESULT 5

## Q8K2T8 MOUSE

ID Q8K2T8 MOUSE PRELIMINARY; PRT; 535 AA.

AC Q8K2T8;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE RIKEN cDNA 5730511K23.

GN Name=5730511K23Rik;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muroidea; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=CZECH II, and C57BL/6;

RC TISSUE=Brain, and

RC Mammary tumor metastatized to lung. MMTV-LTR/Wnt1 model. Expression

RC driven by an MMTV-LTR enhancer.;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CZECH II;
RC TISSUE=Mammary tumor metastatized to lung. MMTV-LTR/Wnt1 model.
RC Expression driven by an MMTV-LTR enhancer.;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Director MGC Project;
RL Submitted (OCT-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC029843; AAH29843.1; -; mRNA.
DR EMBL; BC083337; AAH83337.1; -; mRNA.
DR Ensembl; ENSMUSG00000003437; Mus musculus.
DR MGI; MGI:1923988; 5730511K23Rik.
DR InterPro; IPR007133; Pfam.
DR Pfam; PF03985; Pfam1; 1.
SQ SEQUENCE 535 AA; 60518 MW; 7A5EAB1284988070 CRC64;

```

Db	301	NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Qy	361	LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	420
			:
Db	361	LEAQEARKAQLNHEPEEEEEEMEAEKEAGGSDEEQEGSSSEKEGSEDEHSGSESDR	420
Qy	421	EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN	480
Db	421	EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAHRGSDN	480
Qy	481	DSDSGSNGGGQR----SRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531
Db	481	DSDSGSDGGGQRSRSQRSRSRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD	535

## RESULT 6

09JJ99 MOUSE

ID Q9JJ99 MOUSE PRELIMINARY; PRT; 535 AA.

AC 09JJ99;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Mus musculus brain cDNA, clone MNCb-6444, similar to Homo sapiens cDNA

DE FLJ11123, clone PLACE1006167.

GN      Name=5730511K23Rik;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muroidea; Muridae; Murinae; Mus.

OX NCBI TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL;

RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,

RA Hashimoto K.;

RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AB041615; BAA95098.1; -; mRNA.

DR Ensembl; ENSMUSG00000003437; Mus musculus.

DR MGI; MGI:1923988; 5730511K23Rik.

DR InterPro; IPR007133; Paf1.

DR Pfam; PF03985; Paf1; 1.

SQ SEQUENCE 535 AA; 60534 MW; 6D7EEB1ECDC8C075 CRC64;

Query Match 98.0%; Score 2708; DB 2; Length 535;

Best Local Similarity 97.9%; Pred. No. 1.2e-119;

Matches 524; Conservative 2; Mismatches 5; Indels 4; Gaps 1;

Qy 1 MAPTIQTQAQREDGHRPNShRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfVQ 60

Db 1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRVQ 60

Qy 61 YKATSLEKQHKKHDLLETPDLGVTIDLINPDTYRIDPNVLLDPADEKLLEEIIQAPTSSKR 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLLEEIEIQAPTSSKR 120

[illegible]

Db 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180

Qy 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEM 240  
 |||

Db 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDLPAPKDTSGAAALEM 240

Qy 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300  
 |||

Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360  
 |||

Db 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360

Qy 361 LEAQEARKAQLLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER 420  
 |||

Db 361 LEAQEARKAQLLENHEPEEEEEEMEAEKEAGGSDEEQEGSSSEKEGSEDEHSGSESDR 420

Qy 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480  
 |||

Db 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAHRGSDN 480

Qy 481 DSDSGSNGGGQR----SRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
 |||:|||| |||

Db 481 DSDSGSDGGGQRSRSQRSRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 535

# RESULT 7

## Q4V886\_RAT

ID Q4V886\_RAT PRELIMINARY; PRT; 535 AA.

AC Q4V886;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein RGD1306219\_predicted.

GN Name=RGD1306219\_predicted;

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Placenta;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Placenta;  
 RG NIH MGC Project;  
 RL Submitted (JUN-2005) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC097494; AAH97494.1; -; mRNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 535 AA; 60546 MW; 48432E1DA398806F CRC64;

Query Match 97.9%; Score 2705; DB 2; Length 535;  
 Best Local Similarity 97.8%; Pred. No. 1.6e-119;  
 Matches 523; Conservative 2; Mismatches 6; Indels 4; Gaps 1;

Qy	1	MAPT	IQTQAQREDGHRPNSHRTLPERSGVVC	RVKYCNSLPDIPFDPKFITYPFDQNR	FVQ	60
Db	1	MAPT	IQTQAQREDGHRPNSHRTLPERSGVVC	RVKYCNSLPDIPFDPKFITYPFDQNR	FVQ	60
Qy	61	YKATSLEKQHKHDL	LTEPDLGVTIDLINPD	TYRIDPNVLLDPADEKL	LEEEIQAPTSSKR	120
Db	61	YKATSLEKQHKHDL	LTEPDLGVTIDLINPD	TYRIDPNVLLDPADEKL	LEEEIQAPTSSKR	120
Qy	121	SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQ	QFT	EEEEIYKDRDSQITAIE		180
Db	121	SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQ	QFT	EEEEIYKDRDSQITAIE		180
Qy	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM				240
Db	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM				240
Qy	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKR	QDEEEMDYAPDDVYDYKIAREYNWNVK			300
Db	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKR	QDEEEMDYAPDDVYDYKIAREYNWNVK			300
Qy	301	NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE				360
Db	301	NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE				360
Qy	361	LEAQEARKAQLLENHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER				420
Db	361	LEAQEARKAQLLENHEPEEEEEEMEAEKEAGGSDEEHEKGSSEKEGSEDEHSGSESER				420
Qy	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN				480
Db	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAHRGSDN				480
Qy	481	DSDSGSNGGGQR----	SRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD			531
Db	481	DSDSGSDGGGQRSRSQSR	SRSRASPFPSGSEHSAQEDGSEAAASDSSEADSDSD			535

075239 HUMAN

Query Match 91.6%; Score 2531.5; DB 2; Length 510;  
Best Local Similarity 91.7%; Pred. No. 2.2e-111;  
Matches 498; Conservative 0; Mismatches 0; Indels 45; Gaps 3;

Qy	1	MAPT	IQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNslPDIPFDpKFITYPFdQNRfVQ	60
Db	1	MAPT	IQTQAQREDGH-----RSGVVCrvKYCNslPDIPFDpKFITYPFdQNRfVQ	50
Qy	61	YKATSLEKQHKHDL	LLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIqAPTSSKR	120
Db	51	YKATSLEKQHKHDL	LLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIqAPTSSKR	110
Qy	121	SQQHAKVVPWMrKTEY	ISTEFNryGISNEKPEVKIGVSVKQqFTEEEIyKDRDSQITAIE	180
Db	111	SQQHAKVVPWMrKTEY	ISTEFNryGISNEKPEVKIGVSVKQqFTEEEIyKDRDSQITAIE	170
Qy	181	KTFEDAQKSISQHY	SKPRVTPVEVMPVFPDFKMWINpCAQVIFdSDpAPKDTSGAAALEM	240
Db	171	KTFEDAQKS-----	-----MWINpCAQVIFdSDpAPKDTSGAAALEM	207
Qy	241	MSQAMIRGMMDEEG	NQFVAYFLPVEETLKKRKRdQEEEMdyAPDDVYdyKIAREYNWNVK	300
Db	208	MSQAMIRGMMDEEG	NQFVAYFLPVEETLKKRKRdQEEEMdyAPDDVYdyKIAREYNWNVK	267
Qy	301	NKASKGYEENyFFI	FREGDGVYYNELETR-----VRLSKRRaKAGVQSGTNAL	348

Db 268 NKASKGYEENYFFIFREGDGVYYNELETRYSAHSYLI SLDLVRLSKRRAKAGVQSGTNAL 327

Qy 349 LVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEG 408  
 |||||

Db 328 LVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEG 387

Qy 409 SEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDD 468  
 |||||

Db 388 SEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDD 447

Qy 469 EDRGQAQGGSDNDS DSGSNGGGQRSRSHRSASPFPSPGSEHSAQEDGSEAAASDSSEADS 528  
 |||||

Db 448 EDRGQAQGGSDNDS DSGSNGGGQRSRSHRSASPFPSPGSEHSAQEDGSEAAASDSSEADS 507

Qy 529 DSD 531  
 |||

Db 508 DSD 510

RESULT 9

Q6P2Y1\_XENTR

ID Q6P2Y1\_XENTR PRELIMINARY; PRT; 520 AA.

AC Q6P2Y1;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE Hypothetical protein MGC76249.

GN Name=MGC76249;

OS *Xenopus tropicalis* (Western clawed frog) (*Silurana tropicalis*).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OC Xenopodinae; *Xenopus*; *Silurana*.

OX NCBI\_TaxID=8364;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Embryo;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC064253; AAH64253.1; -; mRNA.  
 DR InterPro; IPR007133; Paf1.  
 DR Pfam; PF03985; Paf1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 520 AA; 59064 MW; 76D526C423C459A7 CRC64;

Query Match 79.0%; Score 2183.5; DB 2; Length 520;  
 Best Local Similarity 80.2%; Pred. No. 5.1e-95;  
 Matches 429; Conservative 40; Mismatches 45; Indels 21; Gaps 8;

Qy	1	MAPT	IQTQAQREDGHR	PNSHRTLPERSGVVC	RVKYCN	SLPDIPFDPKFITY	PFDQNR	FVQ	60
Db	1	MAPT	IQTQAQREDGHR	SSSHRTVPERSG	VVC	RVKYCN	TLPDIPFDPKFITY	PFDQNR	60
Qy	61	YKAT	SLEKQHKHDL	LLETPDLGVTID	LINPDTYRIDP	NVLLDPADEK	LLEEEIQA	PTSSKR	120
Db	61	YKAT	SLEKQHKHDL	LLETPDLGVTID	LINPDTYRIDP	NVTLDDIADEK	LLEEEIQA	PSSSKR	120
Qy	121	SQQH	AKVVPWMRKTEY	ISTEFNRYGIS	NEKPEVKIG	SVKQQFTEEEI	YKDRDSQ	ITAIE	180
Db	121	SQQH	AKVVPWMRKTEY	ISTEFNRYGVS	NEKPEVKIG	SVKQQFTEEDI	YKDRDSQ	ISAIE	180
Qy	181	KTFE	DAQKSISQHY	SKPRVTPVEV	MPVFPDFK	MWINPCAQVI	FSDPAPK	DTSGAA	240
Db	181	KTFE	DAQKPISQHY	SKPRVTPVEV	MPVFPDFK	MWINPCAQVI	FSDPAPK	DASGSA	240
Qy	241	MSQA	MIRGMMDEEG	NQFVAYFLP	VEETLKKR	KRDQEEEM	DYAPDDV	YDYKIA	300
Db	241	MSQA	MIRGMMDEEG	NQFVAYFLP	GREETMRK	RKDQEEGL	DYMPEDI	YDYKIA	300
Qy	301	NKAS	KGYEENYFFI	FREGDGVY	YNELETRV	RLSKRR	AKAGVQ	SGTNALL	360
Db	301	NKAS	KGYEENYFFI	FREGDGVY	YNELETRV	RLSKRR	VKAGVQ	SGTNAV	360
Qy	361	LEAQ	EARKAQL	ENHEPEEEEE	EEMETEEKE	AGGSD-EEQ	EKGSSSE	KEGSEDE	419
Db	361	LEAQ	EARRAQL	ENHEPEEEEE	EIEV---DQ	ETQGS	DAEDGEK	SGSGSEK	417
Qy	420	REEG	RDEASDKS	GSGEDESSE	DEARAARD	KEEIFGSD	ADSEDD	ADSDDED	479
Db	418	REEA	EEEEKEDE	-----E	KESSSEED	RARAARD	KEEIFGSD	---DDSD	466
Qy	480	NDSD	SGSNG---	GGQ	RSRSHRS	SASPFPSG	SEHSAQ	EDGSEAA	531
Db	467	GE-	DSGSDDEE	EKGQGR	RSRSASS	SPF--GSD	HQQENED	QSASDQ	518

# RESULT 10

Q4U0S5\_BRARE

ID Q4U0S5\_BRARE PRELIMINARY; PRT; 503 AA.

AC Q4U0S5;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE PD2-like protein.  
 OS Brachydanio rerio (Zebrafish). (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Amsterdam A., Hopkins N.;  
 RT "Danio rerio PD2-like mRNA."  
 RL Submitted (APR-2005) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; DQ022213; AAY44602.1; -; mRNA.  
 SQ SEQUENCE 503 AA; 58216 MW; B8EE86A45B9D9DEE CRC64;

Query Match 73.4%; Score 2030; DB 2; Length 503;  
 Best Local Similarity 74.2%; Pred. No. 8.2e-88;  
 Matches 402; Conservative 40; Mismatches 50; Indels 50; Gaps 8;

Qy 1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFQDQNRVQ 60  
 |||||::|||:||||| |||||:|||||  
 Db 1 MAPTIQTQAQREDGHRSSAHRTPERSGVVCrvKYGNSLPDIPFDPKFITYPFQDQHRVQ 60

Qy 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDITYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120  
 |||||:|||||:|||||:|||||:|||||  
 Db 61 YKATSLEKQHKHELLTEPD LGVTIDLINPDITYRIDPNILLDPADEKLLEEEIQAPSSKR 120

Qy 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180  
 |||||:||||| |||||:||||| |||||  
 Db 121 SQQHAKVVPWMRKTEYISTEFNRYGVSNEKVEVKIGVSVKQQFTEEEIYKDRDSQIAAIE 180

Qy 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240  
 |||||:|||||:|||||:|||||:||||| ||||| |||||: |||||  
 Db 181 KTFEDAQKSISQHYSKPRVTPVEVLPVFPDFKMWINPCAQVIFDSDPAPKDV SAPAGVDM 240

Qy 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYDYKIAREYNWNVK 300  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||  
 Db 241 MSQAMIRGMMDEEGNQFVAYFLPNEDTMRKRKRQDEEELDYMPEEVYDYKIAREYNWNVK 300

Qy 301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360  
 |||||:|||||:|||||:||||| ||||| |||||: ||||| |||||  
 Db 301 NKASKGYEENYFFIFRDADGVYYNELETRVRLSKRRKAGVQSSSTNAVLVCKHRDMNEKE 360

Qy 361 LEAQEARKAQLNHEPEEEEE-EMETEEKEAGGSDEEQ---EKGSSSEKEGSEDEHSGS 416  
 |||||:|||||: |||||: |||||: |||||: |||||: |||||: |||||  
 Db 361 LEAQEARKAQLNHEPEDEEEELDLEKDMQEDSGEEREKPSDSENSESESESEEEEEERPAD 420

Qy 417 ESEREEDG-----RDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDE 469  
 |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||  
 Db 421 EDEEEEDVESVKKRRERKSSGSESGDD-----RQARDEEEIFGSDDDSEEEEEEEEE 473

Qy 470 DRGQAQGGSDNDSGSGNGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSD 529  
 |||||: |||||: |||||: |||||: |||||: |||||: |||||  
 Db 474 -----GGARRRSNSSSV-----QHSASE-----RASDSSDA-SD 501

Qy 530 SD 531



Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360  
 |||  
 Db 291 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 350

Qy 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGS 394  
 |||  
 Db 351 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGS 384

# RESULT 12

## Q9CS63\_MOUSE

ID Q9CS63\_MOUSE PRELIMINARY; PRT; 377 AA.  
 AC Q9CS63;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched  
 DE library, clone:5730511K23 product:PD2 PROTEIN (HYPOTHETICAL 60.0 kDa  
 DE PROTEIN) homolog (Fragment).  
 GN Name=5730511K23Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK017762; BAB30913.1; -; mRNA.  
 DR Ensembl; ENSMUSG00000003437; Mus musculus.  
 DR MGI; MGI:1923988; 5730511K23Rik.  
 DR InterPro; IPR007133; Paf1.  
 DR Pfam; PF03985; Paf1; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 43836 MW; 4ECE00D2D4EF5CEA CRC64;

Query Match 71.8%; Score 1984; DB 2; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-86;  
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQ 60  
 |||  
 Db 1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQ 60

Qy 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120  
 |||  
 Db 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120

Qy 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGSVKQQFTEEEIYKDRDSQITAIE 180  
 |||  
 Db 121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGSVKQQFTEEEIYKDRDSQITAIE 180

Qy 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240  
 |||  
 Db 181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240

Qy 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYDYKIAREYNWNVK 300  
 |||  
 Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360  
 |||  
 Db 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360

Qy 361 LEAQEARKAQLNHEPE 377  
 |||  
 Db 361 LEAQEARKAQLNHEPE 377

# RESULT 13

Q68F51\_XENLA

ID Q68F51\_XENLA PRELIMINARY; PRT; 407 AA.  
 AC Q68F51;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE LOC446278 protein (Fragment).  
 GN Name=LOC446278;  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative.";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC079993; AAH79993.1; -; mRNA.  
 DR InterPro; IPR007133; Paf1.  
 DR Pfam; PF03985; Paf1; 1.  
 FT NON\_TER 407 407  
 SQ SEQUENCE 407 AA; 47154 MW; 6CE32A7307186F83 CRC64;



Qy 60 QYKATSLEKQHKHDLLEPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSK 119  
 |||:|||||:|||||:|||||:  
 Db 61 QYKATSLEKQHKHDLLEPDLGVTIDLINPDYRIDPSVLLDPADEKLLEEDIQAPSSSK 120

Qy 120 RSQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAI 179  
 |||:|||||:|||||:|||||:  
 Db 121 RSQQHAKVVPWMRKTEYISTEFNRYGVSNEKVEVKIGVSVKQQFTEEEIYKDRDSQISAI 180

Qy 180 EKTFFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALE 239  
 |||:|||||:|||||:|||||:|  
 Db 181 EKTFFEDAQKSITQHYSKPRVTPVEVLVPVFPDFKMWINPCAQVIFDSDPAPKDISGPAGVE 240

Qy 240 MMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNV 299  
 |||:|||||:|||||:|||||:  
 Db 241 MMSQAMIRGMMDEEGNQFVAYFLPNEDTLRKRKRDFEEGVDYMPEDLYDYKIAREYNWNV 300

Qy 300 KNKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEK 359  
 |||:|||||:|||||:|||||:  
 Db 301 KNKASKGYEENYFFIFRDGDGVYNELETRVRLSKRRAKAGTQSTTNAVLVCKHRDMNEK 360

Qy 360 ELEAQ 364  
 ||||  
 Db 361 ELEAQ 365

# RESULT 15

## Q9VN55\_DROME

ID Q9VN55\_DROME PRELIMINARY; PRT; 538 AA.  
 AC Q9VN55;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE CG2503-PA (LD37523p).  
 GN Name=atms; ORFNames=CG2503;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomics perspective.";  
 RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Berkeley Drosophila Genome Project;  
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
 RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirska R., Smith E.,  
 RA Yu C., Rubin G.;  
 RT "Drosophila melanogaster release 4 sequence.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RG FlyBase;  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003605; AAF52095.1; -; Genomic\_DNA.  
 DR EMBL; AY070561; AAL48032.1; -; mRNA.  
 DR Ensembl; CG2503; Drosophila melanogaster.  
 DR FlyBase; FBgn0010750; atms.  
 DR FlyBase; FBgn0010750; CG2503.  
 DR InterPro; IPR007133; Paf1.  
 DR Pfam; PF03985; Paf1; 1.  
 SQ SEQUENCE 538 AA; 60794 MW; D55E95B4F4EB8E51 CRC64;

Query Match 45.0%; Score 1244.5; DB 2; Length 538;  
 Best Local Similarity 50.0%; Pred. No. 7.5e-51;  
 Matches 271; Conservative 66; Mismatches 172; Indels 33; Gaps 11;

Qy 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKfITYPFdQNRfVQ 60  
 | ||| | : | :| :||| | :||| | || : ||| :|||  
 Db 1 MPPTINNSAVNSAAEK-RPQRQTERKSEIICrvKYGNLDPDIPFDLkFLQYPFDShrfVQ 59

Qy 61 YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR 120  
 | ||| : | :||| | ||| :||| : | : | ||| | ||| | ||| |  
 Db 60 YNPTSLERNFKYDVLTEHDLGVTVDLINRELYQADSMTLLDPADEKLLEETLTPTDSVR 119

Qy 121 SQQHAKVVPWMrKTEYISTEFNryGISN-EKPEVKIGVSVKQQFTEEEIYKDRDSQITAI 179  
 | :|| : | | :|| :||| | : | | | :| :|| : | | :|| :||  
 Db 120 SRQHSRTVSWLRKSEYISTEQTRFQPQNLNIEAKVGYNVKKSLREETLYLDREAQIKAI 179

Qy 180 EKTfEDAQKSISQhYskPRVTPVEVMPVFPDFKMWINPCAQVIFdSDPAPKDTSGAAALE 239  
 ||| | : | :||| | ||| :||| | ||| | ||| | : | ||  
 Db 180 EKTfSDTKSEITKhYskPNVVPVEVLPIFPDFTNWKFPCAQVIFdSDPAPAGKNVPAQLE 239

Qy 240 MMSQAMIRGMMDEEGNQFvAYFLPVEETLKKRKRdQEEEMDYAPDDVYdYKIAREYNWNV 299  
 ||| ||| :|| | ||| | :|| :|| | :| :| :||| | |||  
 Db 240 EMSQAMIRGVMDESGEQFvAYFLPTEQTLEKRRTDFINGELYKEEEYeyKIAREYNWNV 299

Qy 300 KNKASKGYEENyFFIFREGDGVYyNELETRVRLSKRRaKAGVQSGTNALLVVKHRDMNEK 359  
 | ||| | ||| : | : ||| | ||| :|| | | | | ||| :|

Db 300 KTKASKGYEENYFFVMRQ-DGIYYNELETRVRLNKRRVKVG-QQPNNTKLVVKHRPLDSM 357  
 Qy 360 ELEAQEARKAQLNHEPEEE-----EEEEM----ETEE-----KEAGGSD----- 395  
 | | |: ||| ||| |||:| |||: : | |:|  
 Db 358 EHRMQRYRERQLEVPGEIIIIVEEVREEEQMQIIIGETEKTSEDA AVGAQAASGADSPAQV 417  
 Qy 396 --EEQEKGSSEKEGSEDEHSGSESEREEDRDEASDKSGSGEDESEDEARAARDKEEI 453  
 : | : | : || ||| | | : : : ||| : | : :  
 Db 418 ARDRQSRSRSTRSGS-SSGSGSGSGSRASSRSKSGSRSGSGSRRTNSPAGSQKSGSR- 475  
 Qy 454 FGSDADSEDDADSDDEDGQAQGGSDNDSDSGS-NGGGQSRSHSRASPFPSGSEHSAQ 512  
 | : | : | | : : : | : | ||| : | | | | ||| || ||| :  
 Db 476 SRSVSRSRSRKSGSRSRSRSRKSGSRSRSGSRSGSGSRSPSRSRSGSPSGSGSSSGSA 535  
 Qy 513 ED 514  
 |  
 Db 536 SD 537

Search completed: April 25, 2006, 09:11:07  
 Job time : 236 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2006, 09:11:27 ; Search time 46 Seconds  
(without alignments)  
954.365 Million cell updates/sec

Title: US-10-721-553-2  
Perfect score: 2764  
Sequence: 1 MAPTIQTQAQREDGHRPNH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2764	100.0	531	1	US-08-933-750C-9	Sequence 9, Appli
2	2764	100.0	531	2	US-09-234-613-9	Sequence 9, Appli
3	2764	100.0	531	2	US-09-647-143-2	Sequence 2, Appli
4	595	21.5	115	2	US-09-513-999C-7407	Sequence 7407, Ap
5	266	9.6	481	2	US-09-248-796A-18683	Sequence 18683, A
6	234.5	8.5	1742	2	US-09-386-962C-4	Sequence 4, Appli
7	234.5	8.5	1742	2	US-09-386-959-4	Sequence 4, Appli
8	233.5	8.4	930	2	US-09-200-650E-3	Sequence 3, Appli
9	226	8.2	918	2	US-09-200-650E-1	Sequence 1, Appli
10	225.5	8.2	933	2	US-08-293-728-2	Sequence 2, Appli
11	225.5	8.2	933	2	US-09-421-868-2	Sequence 2, Appli

12	225.5	8.2	936	2	US-08-956-171E-5249	Sequence 5249, Ap
13	225.5	8.2	936	2	US-08-781-986A-5249	Sequence 5249, Ap
14	223.5	8.1	1315	2	US-09-200-650E-5	Sequence 5, Appli
15	220.5	8.0	1259	2	US-09-949-016-10366	Sequence 10366, A
16	218	7.9	1166	2	US-09-200-650E-7	Sequence 7, Appli
17	212.5	7.7	287	2	US-09-710-279-468	Sequence 468, App
18	212.5	7.7	1092	2	US-09-147-405B-15	Sequence 15, Appl
19	203.5	7.4	414	2	US-09-248-796A-19046	Sequence 19046, A
20	203	7.3	257	2	US-09-461-697-188	Sequence 188, App
21	203	7.3	272	2	US-09-461-697-186	Sequence 186, App
22	199.5	7.2	599	2	US-09-538-092-864	Sequence 864, App
23	198.5	7.2	238	2	US-09-461-697-190	Sequence 190, App
24	196	7.1	781	2	US-09-949-016-9773	Sequence 9773, Ap
25	195	7.1	231	2	US-09-461-697-194	Sequence 194, App
26	195	7.1	232	2	US-09-461-697-192	Sequence 192, App
27	195	7.1	764	2	US-09-538-092-944	Sequence 944, App
28	191	6.9	598	2	US-09-538-092-1083	Sequence 1083, Ap
29	189	6.8	764	2	US-09-370-838-67	Sequence 67, Appl
30	189	6.8	764	2	US-09-854-133-67	Sequence 67, Appl
31	186	6.7	3135	1	US-08-323-170B-2	Sequence 2, Appli
32	186	6.7	3135	2	US-08-954-441-2	Sequence 2, Appli
33	185.5	6.7	1162	1	US-08-728-323A-2	Sequence 2, Appli
34	185.5	6.7	1162	2	US-09-298-568-2	Sequence 2, Appli
35	185.5	6.7	1162	2	US-09-410-399-2	Sequence 2, Appli
36	185.5	6.7	1162	2	US-09-894-273-2	Sequence 2, Appli
37	185	6.7	402	2	US-09-248-796A-18910	Sequence 18910, A
38	184	6.7	674	2	US-08-893-852A-1	Sequence 1, Appli
39	183	6.6	486	2	US-09-710-279-788	Sequence 788, App
40	180	6.5	487	2	US-09-386-962C-14	Sequence 14, Appl
41	180	6.5	487	2	US-09-386-959-65	Sequence 65, Appl
42	178.5	6.5	40	2	US-09-647-143-16	Sequence 16, Appl
43	178.5	6.5	1269	2	US-09-949-016-7349	Sequence 7349, Ap
44	178.5	6.5	1269	2	US-09-949-016-7350	Sequence 7350, Ap
45	178	6.4	1444	2	US-09-949-016-9652	Sequence 9652, Ap

#### ALIGNMENTS

#### RESULT 1

US-08-933-750C-9

; Sequence 9, Application US/08933750C

; Patent No. 5932442

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Bandman, Olga

; APPLICANT: Shah, Purvi

; APPLICANT: Au-Young, Janice

; APPLICANT: Yue, Henry

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

; NUMBER OF SEQUENCES: 98

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

*filed 131*



Db 241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVDYDYKIAREYNWNVK 300

Qy 301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360  
 |||

Db 301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360

Qy 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER 420  
 |||

Db 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER 420

Qy 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480  
 |||

Db 421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480

Qy 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
 |||

Db 481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531

RESULT 2

US-09-234-613-9

; Sequence 9, Application US/09234613

; Patent No. 6132973

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Bandman, Olga

; APPLICANT: Shah, Purvi

; APPLICANT: Au-Young, Janice

; APPLICANT: Yue, Henry

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

; NUMBER OF SEQUENCES: 98

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/234,613

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/933,750

; FILING DATE: September 23, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0356 US

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PITUNOR01
; CLONE: 98974
US-09-234-613-9

```

```

Query Match          100.0%; Score 2764; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 1.4e-227;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MAPTIQTQAQREDGHRPN SHRTLPERSGVVC RVKYCNSLPDIPFDPKFITYPFDQNR FVQ 60
        ||||||||||||||||||
Db      1 MAPTIQTQAQREDGHRPN SHRTLPERSGVVC RVKYCNSLPDIPFDPKFITYPFDQNR FVQ 60

Qy     61 YKATSLEKQHKH DLLTEPDLGVTIDLINPD TYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
        ||||||||||||||||||
Db     61 YKATSLEKQHKH DLLTEPDLGVTIDLINPD TYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120

Qy    121 SQQHAKVVPW MRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFT EEEIYKDRDSQITAI E 180
        ||||||||||||||||||
Db    121 SQQHAKVVPW MRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFT EEEIYKDRDSQITAI E 180

Qy    181 KTFEDAQKSIS QHYSKPRVTPVEVMPVF PDKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
        ||||||||||||||||||
Db    181 KTFEDAQKSIS QHYSKPRVTPVEVMPVF PDKMWINPCAQVIFDSDPAPKDTSGAAALEM 240

Qy    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKR DQEEEMDYAPDDVYDYKIAREYNWNV K 300
        ||||||||||||||||||
Db    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKR DQEEEMDYAPDDVYDYKIAREYNWNV K 300

Qy    301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRR AKAGVQSGTNALLVVKHRDMNEKE 360
        ||||||||||||||||||
Db    301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRR AKAGVQSGTNALLVVKHRDMNEKE 360

Qy    361 LEAQEARKAQL ENHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420
        ||||||||||||||||||
Db    361 LEAQEARKAQL ENHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420

Qy    421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
        ||||||||||||||||||
Db    421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480

Qy    481 DSDSGSNGGGQ RSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531
        ||||||||||||||||||
Db    481 DSDSGSNGGGQ RSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531

```

RESULT 3

US-09-647-143-2  
; Sequence 2, Application US/09647143  
; Patent No. 6680196  
; GENERAL INFORMATION:  
; APPLICANT: Batra, Surinder K.  
; APPLICANT: Hollingsworth, Michael A.  
; APPLICANT: University of Nebraska Board of Regents  
; TITLE OF INVENTION: No. 6680196el Gene That is Amplified and  
; TITLE OF INVENTION: Overexpressed in Cancer and Methods of Use Thereof  
; FILE REFERENCE: UNMC63121  
; CURRENT APPLICATION NUMBER: US/09/647,143  
; CURRENT FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US99/06633  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/079,649  
; PRIOR FILING DATE: 1998-03-27  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 531  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-647-143-2

Query Match 100.0%; Score 2764; DB 2; Length 531;  
Best Local Similarity 100.0%; Pred. No. 1.4e-227;  
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAPT	100.0%	Score 2764	DB 2	Length 531
Db	1	MAPT	100.0%	Pred. No. 1.4e-227		
Qy	61	YKATSLEKQHKHDLLETPDLGVTIDLINPD	100.0%	Score 2764	DB 2	Length 531
Db	61	YKATSLEKQHKHDLLETPDLGVTIDLINPD	100.0%	Pred. No. 1.4e-227		
Qy	121	SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFT	100.0%	Score 2764	DB 2	Length 531
Db	121	SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFT	100.0%	Pred. No. 1.4e-227		
Qy	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	100.0%	Score 2764	DB 2	Length 531
Db	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	100.0%	Pred. No. 1.4e-227		
Qy	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK	100.0%	Score 2764	DB 2	Length 531
Db	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK	100.0%	Pred. No. 1.4e-227		
Qy	301	NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE	100.0%	Score 2764	DB 2	Length 531
Db	301	NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE	100.0%	Pred. No. 1.4e-227		
Qy	361	LEAQEARKAQLLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	100.0%	Score 2764	DB 2	Length 531
Db	361	LEAQEARKAQLLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	100.0%	Pred. No. 1.4e-227		
Qy	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDGQAQGGSDN	100.0%	Score 2764	DB 2	Length 531
Db	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDGQAQGGSDN	100.0%	Pred. No. 1.4e-227		

```

Db          421 EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
Qy          481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531
Db          481 DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531

```

RESULT 4

US-09-513-999C-7407

; Sequence 7407, Application US/09513999C

; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 7407

; LENGTH: 115

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: 25

; OTHER INFORMATION: Xaa=Glu or Gly

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: 26

; OTHER INFORMATION: Xaa=Arg or Ser

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: 110

; OTHER INFORMATION: Xaa=Glu or Gly

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: 114

; OTHER INFORMATION: Xaa=Ala or Gly

US-09-513-999C-7407

Query Match 21.5%; Score 595; DB 2; Length 115;

Best Local Similarity 96.5%; Pred. No. 2.8e-43;

Matches 111; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

Qy          1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNslPDIPFDPKFITYPFdQNRfVQ 60
Db          1 MAPTIQTQAQREDGHRPNSHRTLpXXSGVVCrvKYCNslPDIPFDPKFITYPFdQNRfVQ 60
Qy          61 YKATSLEKQHKHDLlTEPDlGVtIDlINPDtYRIDPNVllDPADEKllEEElQAP 115

```

Db

61 YKATSLEKQHKHDLLETPDLGVTIDLINPDITYRIDPNVLLDPADEKLLEXEIQXP 115

## RESULT 5

US-09-248-796A-18683

; Sequence 18683, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 18683

; LENGTH: 481

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-18683

Query Match 9.6%; Score 266; DB 2; Length 481;

Best Local Similarity 22.0%; Pred. No. 2.8e-14;

Matches 110; Conservative 91; Mismatches 156; Indels 142; Gaps 20;

```

Qy      18 NSHRTL-PERSGVVCRVKYCNSLPDIPFDPKFITY----PFDQNRVQYKATSL-EKQHK 71
          :|:::| | | : :|:| |::| | :||| | | : :| :|| |::
Db      16 SSNKS LKPIRQDYIAKVRYTNNLPPLNPKFIEYNTTDPISTQQEGEYLISSLF RKENF 75

Qy      72 HDLLTEPD--LGVTIDLINPDITY-----RIDPN---VLLDPADEKLLEE 110
          :|: | | |:::| | : : : : : : : : : : : | | | :|:
Db      76 QNLMERIDDLGLDLN LNNRGFLSEDKMNESVGKLYNQ LHPNDRALLRDAGIGKILKN 135

Qy     111 EIQAPTSSKRSQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIG-----VSVKQ 161
          | : | : :|:| | | : :| | | : :| :
Db     136 EPE-----VSFLRRTEYIS-----DRPLSKGGNNLNTATEEIKVKE 171

Qy     162 QFTEEEIYKDRDSQITAEKTFEDAQKSI-----SQHYSKPRVTPVEVMPVFPDFKMWIN 216
          : :::| : | | | : :|:| | :| : | : | :| | | |
Db     172 KLSKDEHF-DADSQLQNVESFTVANESLYDLKNIKHPKKKHLRAVNTWPLLPDTSMLDN 230

Qy     217 PCAQVIF-DSDPAPKDTSGAAALEMMSQAMIRGMMDE-----EGNQFV 258
          : | | : : : : : | | | : :
Db     231 VFINLRFMGSASINRELNNLKKQQQQQQQQNDKKFDEKLFDRALLESSLFKPIKLEGGEWI 290

Qy     259 AYFL--PVEETLKKRKRQEEEMDYAPDDVYD--YKIAREYNWNV-----KNKA 303
          : :| : : | |:: :|:| : :| | : |
Db     291 SMYLLDATNTSTTANDNDNEEQI----NDLYEKLHSTKKEQPINLLDEDEESLETYKFKY 346

Qy     304 SKGYEENYFFIFREGDGV-----YYNELETRVRLSKRRA- 337
          :| | : | : : : : : : : : : :
Db     347 TKNYDMTYQPFHEHENEELAIKFVSDEIEDPVSKDNFKRKRKMAYYYYPINGKIELKKHRAS 406

```

```

QY      338 -----KAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEE 388
          |      | | :|      | : : || : : | | | :|| | ||
Db      407 TNSEINKFIKERTYDGINFIL---REPSTNELKRLDTIRSEYDPMEYEGEDEEEEEEEE 462

QY      389 KEAGGSDEEQEKGSSSEKE 407
          :|      :| :| : :      : : |
Db      463 EEEPLEEEQQQQEVETKEE 481

```

Qy	282	-----APDDVYD---YKIAR-----EYNWNVKNK-----ASKGY-----	307
		:     : :	
Db	1004	GITTTGIINGADNLTIDSGFYKTPKYSVG DYVWEDTNKDG IQDDNEKGISGVKVTLKDEK	1063
Qy	308	-----EENYFFIFREGD-GVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMN	357
		:     :         : : :   : :	
Db	1064	GNIISTTTTIDENGKYQFDNLD SGNYIIHF EKPEGMTQT TANSG-----NDD	1109
Qy	358	EKELEAQEARKA-----QLENHEPEEEEEEEEMETEEKEAGGSDEEQEGKSSSEKEGSE	410
		: : :   : : : : : :     : :   : :	
Db	1110	EKDADGEDVRVTITDHDDFSIDNGYFDDSDSDSDADSDSDSDSDADSDADSDADSDADSDA	1169
Qy	411	DEHSGSESERE-EGDRDEASDK-SGSGEDESSEDEARAARDKEEIFGSDADSEDDADSD	468
		: : :           : : :   :     :   :	
Db	1170	SDSDSDSDSDADSDSDSDSDSDSDADSDSDSDSDSDADSDSDSDSDSDSDSDSDSDSDS	1229
Qy	469	EDRGQAQGGSDNDSGSGNGGGQSRSHSRASPFPSGSEHSAQEDGSEAAASDS-SEAD	527
		: :     :     :   :   :   :   :	
Db	1230	SDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDADSDSDADSDSDSDSDADSDSDSDSDSDAD	1286
Qy	528	SDSD 531	
Db	1287	SDSD 1290	

Db 774 DGERDGTRKVVAKGTINNADNMTVDTFGYLTPKYNVGDYVWEDTNKDG IQDDNEKGISG 833

Qy 157 VSV-----KQQFT--EEEIY-----KDRDSQ 175  
| | :|| | | :||

Db 834 VKVTILKNKNGDTIGTTTTDSNGKYEFTGLENGDYTIEFETPEGYTPKQNSGSDEGKDSN 893

Qy 176 ITAIEKTFEDA-QKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDP----- 227  
| | :|| | : | | : : | |

Db 894 GTKTTTVTKDADNKTIDSGFYKPTYN-----LGDY-VWEDTNKDG IQDDSEKGISGVK 945

Qy 228 -APKDTSGAA----ALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRDOEEEMDY- 281  
|| :| | : : | | : | | : :|

Db 946 VTLKDKNGNAIGTTTTDASGHYQFKGL--ENGSYTVEFETPSGYTPTKANSQDITVDSN 1003

Qy 282 -----APDDVYD---YKIAR----EYNWNVKNK-----ASKGY----- 307  
| : | || : :| | ||

Db 1004 GITTTGIINGADNLTIDSGFYKTPKYSVG DYVWEDTNKDG IQDDNEKGISGVKVTLKDEK 1063

Qy 308 -----EENYFFIFREGD-GVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMN 357  
: || : | | | | : : | : | : :

Db 1064 GNIISTTTTIDENGKYQFDNLDSGNYIIHFEEKPEGMTQTTANS-----NDD 1109

Qy 358 EKELEAQEARKA-----QLENHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGSE 410  
|| : : : | : | : : : : : || : : | : :

Db 1110 EKDADGEDVRVTITDHDDFSIDNGYFDDSDSDSDADSDSDSDSDADSDSDADSDSDA 1169

Qy 411 DEHSGSESERE-EGDRDEASDK-SGSGEDESEDEARAARDKEEIFGSDADSEDDADSD 468  
| | | : : : | | || | | | : : : | : || : || : ||

Db 1170 DSDSDSDSDSDADSDSDSDSDSDSDSDSDADSDSDSDSDSDADSDSDSDSDSDSDSDS 1229

Qy 469 EDRGQAQGGSDNDSGSGNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDS-SEAD 527  
: : || : || | : | | : | | | : : | : || | : ||

Db 1230 DSDSDSDSDSDSDSDSDSDSDSDSDSDSDADSDSDSDSDSDADSDSDSDSDSDAD 1286

Qy 528 SDSD 531  
||||

Db 1287 SDSD 1290

# RESULT 8

US-09-200-650E-3

; Sequence 3, Application US/09200650E

; Patent No. 6680195

## ; GENERAL INFORMATION:

; APPLICANT: Patti, Joseph M.

; APPLICANT: Foster, Timothy J.

; APPLICANT: Hook, Magnus A.O.

; APPLICANT: Eidhinn, Deirdre Ni

; APPLICANT: Perkins, Samuel L.

; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus

; FILE REFERENCE: P06283US2/BAS

; CURRENT APPLICATION NUMBER: US/09/200,650E

; CURRENT FILING DATE: 1998-11-25

; PRIOR APPLICATION NUMBER: 60/066,815

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: 60/098,427



RESULT 9  
 US-09-200-650E-1  
 ; Sequence 1, Application US/09200650E  
 ; Patent No. 6680195  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Patti, Joseph M.  
 ; APPLICANT: Foster, Timothy J.  
 ; APPLICANT: Hook, Magnus A.O.  
 ; APPLICANT: Eidhinn, Deirdre Ni  
 ; APPLICANT: Perkins, Samuel L.  
 ; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus  
 ; FILE REFERENCE: P06283US2/BAS  
 ; CURRENT APPLICATION NUMBER: US/09/200,650E  
 ; CURRENT FILING DATE: 1998-11-25  
 ; PRIOR APPLICATION NUMBER: 60/066,815  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: 60/098,427  
 ; PRIOR FILING DATE: 1998-08-31  
 ; NUMBER OF SEQ ID NOS: 23  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 918  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-09-200-650E-1

Query Match 8.2%; Score 226; DB 2; Length 918;  
 Best Local Similarity 22.6%; Pred. No. 1.8e-10;  
 Matches 125; Conservative 75; Mismatches 208; Indels 146; Gaps 23;

Qy	33	VKYCNSLPDIPF-DPKFITYPFDQNRVQYKATSLEKQHKHDLLEPDLGVTIDLIN---	88
		:                     :	
Db	276	VDYSNSNNTMPIADIK-----STNGDVVAKAT-----YDILTKTYTFVFTDYVNNKE	322
Qy	89	-----PD TYRIDPNVLLDPADEKLLEE-----EIQAPTSSKRS	121
		:       :         :	
Db	323	NINGQFSLPLFTDRAKAPKSGTYDANINI--ADEMFNNKITYNYSSPIAGIDKPNGANIS	380
Qy	122	QQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE-	180
		:             : : :   : : : :	
Db	381	SQIIGVDTASGQNTYKQTVF-----VNPQQRVLGNTWVYIKGYQDKI-EESSGKVSATDT	434
Qy	181	--KTFE--DAQKSISQHYSKPRVTPV-EVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGA	235
		:         :   :       : :	
Db	435	KLRIFEVNDTSKLSDSYADPNDSNLKEVTDQFKNRIYYEHPNVA SIKFGD-----	485
Qy	236	AALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQEEEMDYAPDDVYDYKIAREY	295
		: : :           :   : :	
Db	486	--ITKTYVVLVEGHYDNTG-----KNLKTQVIQENVDPVTNRDYSI---F	525
Qy	296	NWNVKNKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRD	355
		:   :               :	
Db	526	GWNNEN-----VVRYG-----GGSADGDSA-----	545
Qy	356	MNEKE-----LEAQEARKAQLE-----NHEPEEEEEEMETEEKEAGGSDEEQEK	400

```

      :| |:      :: : :   : |           : |||   : : ::   ||| :
Db      546 VNPkdPTPGPPVDPEPSDPDEPEPTPDPEPSDPDEPEPSDPDPDPDSDSDSDSGSDSDSGS 605
Qy      401 GSSSEKEGSEDEHSGSESERE-EGDRDEASDK-SGSGEDESEDEARAARDKEEIFGSDA 458
      | || :   | | |:|: : | | | || | | | |: : : : | :   ||:
Db      606 DSDSESDSDSDSDSDSDSDSDSESDSDSESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 665
Qy      459 DSEDDADSDDEDRGQAQGGSDNDSDSGSGGGQSRSHSRASPFPSGSEHSAQEDGSEA 518
      ||: ::||| |   ::   ||:|||| |:   | | | | | | |: : |
Db      666 DSDSESDSDSESDSESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 724
Qy      519 AASDS-SEADSDSD 531
      : ||| |:|||||
Db      725 SDSDSDSDSDSDS 738

```

RESULT 10

US-08-293-728-2

; Sequence 2, Application US/08293728D

; Patent No. 6008341

; GENERAL INFORMATION:

; APPLICANT: Foster, Timothy J.

; APPLICANT: McDevitt, Damien L.

; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene

; FILE REFERENCE: 05344.105011

; CURRENT APPLICATION NUMBER: US/08/293,728D

; CURRENT FILING DATE: 1994-08-22

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 933

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-08-293-728-2

```

Query Match          8.2%;   Score 225.5;   DB 2;   Length 933;
Best Local Similarity 21.4%;   Pred. No. 2.1e-10;
Matches 103;   Conservative 76;   Mismatches 187;   Indels 115;   Gaps 18;

```

```

Qy      83 TIDLINP--DTYRIDPNVLLDPADEKLLLEEIIQAPTSSKRSQQHAKVVPWMRKTEYISTE 140
      ||| |:   :|||   : ::|: : ::   || :
Db      383 TIDQIDKTNN TYR--QTIYVNPSGDNVI-----APVLT----- 413
Qy      141 FNRYGISNEKPEVKIGVSVKQQFTEEEIYK-----DRDSQITAIEKTFEDAQKSIHQHYS 195
      | ||   : || | ::||   |   : |||   |:: :
Db      414 -----GNLKPNTDSNALIDQQNTSIKVYKVDNAADLSESYFVNPNFEDVTNSVNITFP 467
Qy      196 KPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEMM-----SQAMIRGM- 249
      |   ||   || ::   |   ||   :: | ||   | : | |
Db      468 NPNQYKVEFNT--PDDQITTPYIVVNGHIDP---NSKGD LALRSTLYGYNSNIWRSMS 522
Qy      250 MDEEGNQFVAYFL-----PVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWNVKNKA 303
      | |   ||:   :: : : |: |:: |:: : :
Db      523 WDNE----VAFNNGSGSGDGIDKPVVPEQPDEPGEIEPIPED-----SDS 563
Qy      304 SKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALL-----VVK 352
      | :           | : : | ||::

```

Db	564	DPGSDSG-----SDSNSDSGSDSGSDSTSDSGSDSASDSDSAS	601
Qy	353	HRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQKGSSEKEGESEDE	412
Db	602	: :   : : : : : : :    : :	661
Qy	413	HSGSESERE-EGDRDEASDKSGSGEDESSSEDEARAARDKEEIFGSDADSED DADSDDedr	471
Db	662	: : : :            : : :   :   : :  :    :	720
Qy	472	GQAQGGSNDNSDSGNGGGQRSRSHRSASPFPSPGSEHSAQEDGSEAAASDS-SEADSDS	530
Db	721	:   :      :            : :   :     :	779
Qy	531	D 531	
Db	780	D 780	

US-09-421-868-2

; GENERAL INFORMATION:

; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene

; CURRENT APPLICATION NUMBER: US/09/421,868

: PRIOR APPLICATION NUMBER: 08/293,728

; PRIOR FILING DATE: 1994-08-22

: NUMBER OF SEQ ID NOS: 20

```
; SOFTWARE: PatentIn Ver. 2.0
```

; SEO ID NO 2

```

; LENGTH: 933

```

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-421-868-2

Best Local Similarity 21.4%; Pred. No. 2.1e-10;

Best Local Similarity 21.4%; Fied. No. 2715 10;  
Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;

Qy 83 TIDLNP--DTYRIDPNVLLDPADEKLLLEEIQAPTSSKRSQQHAKVVPWMRKTEYISTE 140  
||| |: :||| : ::| : :: || :

QY 141 FNRYGISNEKPEVKIGSVKQQFTEEEIYK-----DRDSQITAIKTFEDAQKSISQHYS 195  
 | | : | | : | | : | | | : : |

Qy 196 KPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEMM-----SQAMIRGM- 249  
| | | | :: | | :: | | | : | |

Qy 250 MDEEGNQFVAYFL-----PVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWNVKNKA 303



```

;          FILING DATE: January 3, 1997
;
;  ATTORNEY/AGENT INFORMATION:
;
;          NAME: Mark J. Hyman
;
;          REGISTRATION NUMBER: 46,789
;
;          REFERENCE/DOCKET NUMBER: PB248P1
;
;  TELECOMMUNICATION INFORMATION:
;
;          TELEPHONE: (240) 314-1224
;
;          TELEFAX: (301) 309-8439
;
;  INFORMATION FOR SEQ ID NO: 5249:
;
;    SEQUENCE CHARACTERISTICS:
;
;      LENGTH: 936 amino acids
;
;      TYPE: amino acid
;
;      STRANDEDNESS: single
;
;      TOPOLOGY: linear
;
;    MOLECULE TYPE: protein
;
;    SEQUENCE DESCRIPTION: SEQ ID NO: 5249:
US-08-956-171E-5249

```

```

Query Match          8.2%;  Score 225.5;  DB 2;  Length 936;
Best Local Similarity 21.4%;  Pred. No. 2.1e-10;
Matches 103;  Conservative 76;  Mismatches 187;  Indels 115;  Gaps 18;

```

```

Qy      83 TIDLINP--DTYRIDPNVLLDPADEKLLLEEEIQAPTSSKRSQQHAKVVPWMRKTEYISTE 140
      ||| |:  :|||  : ::| : : :  ||  :
Db      392 TIDQIDKTNNNTYR--QTIYVNPSPGDNVI-----APVLT----- 422

Qy      141 FNRYGISNEKPEVKIGVSVKQQFTEEEIYK-----DRDSQITAIEKTFEDAQKSISQHY 195
      | ||  : || | :|||  |  : |||  |:: :
Db      423 -----GNLKPNTDSNALIDQQNTSIKVYKVDNAADLSESYFVNPNFEDVTNSVNI 476

Qy      196 KPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEMM-----SQAMIRGM- 249
      |  ||  || ::  |  ||  :: | ||  : || |
Db      477 NPNQYKVEFNT--PDDQITTPYIVVVNGHIDP---NSKGD LALRSTLYGYNSNIIWRSMS 531

Qy      250 MDEEGNQFVAYFL-----PVEETLKKRKRQEEEMDYAPDDVYDYKIAREYNWNVKNKA 303
      | |  ||:  :: :  : | : |:: | :  : :
Db      532 WDNE---VAFNNGSGSGDGIDKPVVPEQPDEPGEIEPIPED-----SDS 572

Qy      304 SKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALL-----VVK 352
      | :  |  : : | ||::
Db      573 DPGSDSG-----SDSNSDSGSDSGSDSTSDSGSDSASDSDSAS 610

Qy      353 HRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDE 412
      |  | :: |  : : : : : : : : || : : | : : |
Db      611 DSDSASDSDSASDSDSASDSDSDNDS DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 670

Qy      413 HSGSESERE-EGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDED 471
      | |:: : : | | || | | | : : : | : ||::| : ||| :
Db      671 DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 729

Qy      472 GQAQGGSDNDS DSGSNGGGQRSRSHRSASPFPSGSEHSAQEDGSEAAASDS-SEADSDS 530
      :  ||:|||| | :  | | | | | | : : |  : |||| |::|||
Db      730 SDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSASDSDSDSDSDSDS 788

Qy      531 D 531
      |
Db      789 D 789

```

RESULT 13

US-08-781-986A-5249

; Sequence 5249, Application US/08781986A

; Patent No. 6737248

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 5255

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/781,986A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Benson, Bob

; REGISTRATION NUMBER: 30,446

; REFERENCE/DOCKET NUMBER: PB248PP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 5249:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 936 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-781-986A-5249

Query Match 8.2%; Score 225.5; DB 2; Length 936;

Best Local Similarity 21.4%; Pred. No. 2.1e-10;

Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;

```

Qy      83 TIDLINP--DTYRIDPNVLLDPADEKLLEEEIQAPTSSKRSQQHAKVVPWMRKTEYISTE 140
      ||| |: :||| : ::| : : : || :
Db      392 TIDQIDKTNNTYR--QTIYVNPSPGDNVI-----APVLT----- 422

Qy      141 FNRYGISNEKPEVKIGVSVKQQFTTEEEIYK-----DRDSQITAIEKTFEDAQKSISQHY 195
      | || : ||| : ::| | : ||| : ||| |:: :
Db      423 -----GNLKPNTDSNALIDQQNTSIKVYKVDNAADLSESYFVNPENFEDVTNSVNITFP 476

```

[illegible]

RESULT 14

US-09-200-650E-5

; Sequence 5, Application US/09200650E

; Patent No. 6680195

## ; GENERAL INFORMATION:

; APPLICANT: Patti, Joseph M.

; APPLICANT: Foster, Timothy J.

; APPLICANT: Hook, Magnus A.O.

; APPLICANT: Eidhinn, Deirdre Ni

; APPLICANT: Perkins, Samuel L.

; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus

; FILE REFERENCE: P06283US2/BAS

; CURRENT APPLICATION NUMBER: US/09/200,650E

; CURRENT FILING DATE: 1998-11-25

; PRIOR APPLICATION NUMBER: 60/066,815

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: 60/098,427

; PRIOR FILING DATE: 1998-08-31

```
; NUMBER OF SEQ ID NOS: 23
```

```
; SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 5

; LENGTH: 1315

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-200-650E-5

## Query Match

8.1%; Score 223.5; DB 2; Length 1315;

Best Local Similarity 21.7%; Pred. No. 5.1e-10;  
Matches 121; Conservative 85; Mismatches 225; Indels 127; Gaps 23;

```
Qy      27  SGVVCRVKYCNSLPDIPFDPKFITYPFDQNRVQYKATSLEKQHKHDLLETPDLGVTIDL 86
      :||:      |  ::  |  |  |  :      |  |  |  :  |||: |
Db      772  TGV-----NGADNMTLDSGF--YKTPKYNLGNVWEDTNKDGKQDSTEKGISGVTVTL 823

Qy      87  INPD-----TYRID-----PNVLLDPADEKLLEEEIQAP 115
      | :      ||::      | :  ||  ::  :
Db      824  KNEGEVLQTTKTKDKGKYQFTGLENGTYKVEFETPSGYTPTQVSGTDEG-IDSNGTST 882

Qy     116  TSSKRSQQHAKV-----VPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYK 170
      |  : : :  :  |  :|:  :|:  :|:  :  :  ||:|  |
Db      883  TGVIKDKDNDTIDSGFYKPTYNLGDYVWEDTNKNGVQDKDEKGISGVTV-----TLK 934

Qy     171  DRDSQITAIKTFEDAQKSISQ----HYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSD 226
      | : ::  | |:  :  :  |  ||  |  :  :
Db      935  DENDKVLKTVTTDENGKYQFTDLNNGTYKVEFETPSGYTPT-----SVTSGN 981

Qy     227  PAPKDTSGAAALEMMSQAMIRGMMDEEG-----NQFVAYFLPVEETLKKRKRQEE 277
      ||::|  ::  |  |  :  |  :|  |  :|  |  :|  |
Db      982  DTEKDSNGLTTTGVIKDA--DNMTLDSGFYKTPKYSLGDYVWY-----DSNKDGKQDSTE 1034

Qy     278  EMDYAPDDVYDYKIAREYNWNVKNK--ASKGYEENYFFIFREGDGVYNELETRVRLSKR 335
      :  :  | |:  | | :  :  :||  :  |  |  |  :|  |
Db     1035  K-----GIKDVKVTL---LNEKGEVIGTTKTDENGKYCFDNLDSGKY-----KVIFEK- 1079

Qy     336  RAKAGV-QSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGS 394
      ||:  ||||  |  |  | :  :  :|  ||  :  :
Db     1080  --PAGLTQTGTNTTEDDKDADGGEVDVTITDHDDFTLDNGYEEETSD-----S 1126

Qy     395  DEEQEKGSSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEEIF 454
      | : :  | |::  |  | |:|:  :|  | ||  |  |  | : : :  | :
Db     1127  DSDSDSDSDSDRSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 1184

Qy     455  GSDADSEDDADSDDEDRGQAQGGSDNDSDSGSGNGGGQSRSHRSASPFPSGSEHSAQED 514
      ||:||:  |:|||  :  :  ||:||||  |  :  |  |  |  |  | :  :  |
Db     1185  DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 1237

Qy     515  GSEAAASDS-SEADSDSD 531
      :  |||  |::| |||
Db     1238  SDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 1255
```

RESULT 15

US-09-949-016-10366

; Sequence 10366, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10366
; LENGTH: 1259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10366
```

```
Query Match          8.0%; Score 220.5; DB 2; Length 1259;
Best Local Similarity 23.1%; Pred. No. 8.6e-10;
Matches 93; Conservative 57; Mismatches 160; Indels 93; Gaps 16;

Qy      171 DRDSQITAIEKTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPK 230
        | || |           | | : | |           : | : | | | |
Db      297 DHDSSI-----GQNSDSKEYYDPEGKE-----DPHNEV--DGDKTSK 331

Qy      231 DTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLK-----KR-----KRDQEEE 278
        : |           | : : | : |           : | : | | | :
Db      332 SEENSA-----GIPEDNGSQ-----RIEDTQQLNHRESKRVENRITKESETHA 374

Qy      279 MDYAPDDVYDYKIAREYNWNVKNKASKGYE-----ENYFFIFREGDGVYYNE---LETRV 330
        : : | : | | | : : || |           : | : | : | : |
Db      375 VGKSQDKGIEIKGPSSGNRNITKEVGKGNEGKEDKGQHGMILGKGNVKTQGEVVNIEGPG 434

Qy      331 RLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEE-----E 382
        : | : | | | : | | : : : : : | : |
Db      435 QKSEPGNKVG-HSNTGS-----DSNSDGYDSYDFDDKSMQGD PNSSDESNGNDANS 486

Qy      383 EMETEEKEAG----GSDEEQEKGSSSEKEGSEDEHSGSESEREEGDR-----DEASDK 431
        | : | | | || : : | : : || : | | : | : : ||
Db      487 ESDNNSSSRGDASYNSDESKDNGNGSDSKGAEDDDSDSTSDTNNSDSNGNGNNGNDNDK 546

Qy      432 SGSGE---DESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDNDSDSGSNG 488
        | || : | | | : : : : || : | | | : : || | :
Db      547 SDSGKGKSDSSSDSDSSSDSSSDSDSSSDSDSSSDSNSSSDSDSSSDSDSSSDSDS-SDS 605

Qy      489 GGQRSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531
        | | : | | : | : | : || : || |
Db      606 SNSSDSSDSSSDSDSSSDSDSSSDSKSDSSKSESDSDSDSKSD 648
```

```
Search completed: April 25, 2006, 09:12:43
Job time : 48 secs
```

OM protein - protein search, using sw model

Run on: April 25, 2006, 09:22:24 ; Search time 164 Seconds  
 (without alignments)  
 1352.850 Million cell updates/sec

Title: US-10-721-553-2  
 Perfect score: 2764  
 Sequence: 1 MAPTIQTQAQREDGHRPNSH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2764	100.0	531	3	US-09-840-787-9	Sequence 9, Appli
2	2764	100.0	531	5	US-10-721-553-2	Sequence 2, Appli
3	2658.5	96.2	553	5	US-10-450-763-50041	Sequence 50041, A
4	2464	89.1	473	3	US-09-986-480-410	Sequence 410, App
5	2464	89.1	473	5	US-10-863-332-410	Sequence 410, App
6	1244.5	45.0	538	6	US-11-097-143-4281	Sequence 4281, A
7	622	22.5	133	5	US-10-450-763-50040	Sequence 50040, A
8	452	16.4	86	4	US-10-424-599-223174	Sequence 223174,
9	375.5	13.6	571	4	US-10-424-599-213359	Sequence 213359,
10	337.5	12.2	644	4	US-10-437-963-116147	Sequence 116147,
11	283	10.2	475	5	US-10-450-763-49771	Sequence 49771, A

12	238.5	8.6	286	4	US-10-425-115-202470	Sequence 202470,
13	237.5	8.6	1633	4	US-10-282-122A-70437	Sequence 70437, A
14	234.5	8.5	1742	4	US-10-615-383-4	Sequence 4, Appli
15	234.5	8.5	1742	4	US-10-690-184-4	Sequence 4, Appli
16	234.5	8.5	1742	4	US-10-689-082-4	Sequence 4, Appli
17	234	8.5	953	4	US-10-282-122A-44457	Sequence 44457, A
18	233.5	8.4	930	5	US-10-744-672-3	Sequence 3, Appli
19	233.5	8.4	930	5	US-10-744-616-3	Sequence 3, Appli
20	233.5	8.4	947	5	US-10-470-048B-86	Sequence 86, Appl
21	233	8.4	932	3	US-09-815-242-5578	Sequence 5578, Ap
22	233	8.4	932	3	US-09-815-242-12438	Sequence 12438, A
23	230.5	8.3	1385	4	US-10-282-122A-44324	Sequence 44324, A
24	230	8.3	877	4	US-10-282-122A-70428	Sequence 70428, A
25	229	8.3	1920	4	US-10-282-122A-71413	Sequence 71413, A
26	228	8.2	428	4	US-10-451-467A-448	Sequence 448, App
27	226	8.2	913	5	US-10-470-048B-63	Sequence 63, Appl
28	226	8.2	918	5	US-10-744-672-1	Sequence 1, Appli
29	226	8.2	918	5	US-10-744-616-1	Sequence 1, Appli
30	225.5	8.2	933	5	US-10-470-048B-93	Sequence 93, Appl
31	225.5	8.2	936	2	US-08-781-986A-5249	Sequence 5249, Ap
32	225.5	8.2	936	4	US-10-329-624-5249	Sequence 5249, Ap
33	224.5	8.1	194	5	US-10-450-763-41624	Sequence 41624, A
34	224	8.1	265	4	US-10-073-256-78	Sequence 78, Appl
35	223.5	8.1	1315	5	US-10-744-672-5	Sequence 5, Appli
36	223.5	8.1	1315	5	US-10-744-616-5	Sequence 5, Appli
37	223.5	8.1	1315	5	US-10-470-048B-124	Sequence 124, App
38	221.5	8.0	1349	3	US-09-815-242-5898	Sequence 5898, Ap
39	221.5	8.0	1349	3	US-09-815-242-13137	Sequence 13137, A
40	220.5	8.0	1253	4	US-10-363-798-2	Sequence 2, Appli
41	220	8.0	717	5	US-10-470-048B-58	Sequence 58, Appl
42	219	7.9	1021	3	US-09-815-242-5471	Sequence 5471, Ap
43	219	7.9	1021	3	US-09-815-242-12544	Sequence 12544, A
44	218	7.9	406	4	US-10-451-467A-262	Sequence 262, App
45	218	7.9	1166	5	US-10-744-672-7	Sequence 7, Appli

# ALIGNMENTS

## RESULT 1

US-09-840-787-9

; Sequence 9, Application US/09840787

; Patent No. US20020058264A1

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti  
; Hillman, Jennifer L.  
; Bandman, Olga  
; Shah, Purvi  
; Au-Young, Janice  
; Yue, Henry  
; Guegler, Karl J.  
; Corley, Neil C.

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

; NUMBER OF SEQUENCES: 98

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive

```

;          CITY: Palo Alto
;          STATE: CA
;          COUNTRY: USA
;          ZIP: 94304
;
;  COMPUTER READABLE FORM:
;          MEDIUM TYPE: Diskette
;          COMPUTER: IBM Compatible
;          OPERATING SYSTEM: DOS
;          SOFTWARE: FastSEQ for Windows Version 2.0
;
;  CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/09/840,787
;          FILING DATE: 23-Apr-2001
;          CLASSIFICATION: <Unknown>
;
;  PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: 09/518,865
;          FILING DATE: <Unknown>
;
;  ATTORNEY/AGENT INFORMATION:
;          NAME: Billings, Lucy J.
;          REGISTRATION NUMBER: 36,749
;          REFERENCE/DOCKET NUMBER: PF-0356 US
;
;  TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 415-855-0555
;          TELEFAX: 415-845-4166
;          TELEX: <Unknown>
;
;  INFORMATION FOR SEQ ID NO: 9:
;          SEQUENCE CHARACTERISTICS:
;          LENGTH: 531 amino acids
;          TYPE: amino acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;
;  IMMEDIATE SOURCE:
;          LIBRARY: PITUNOR01
;          CLONE: 98974
;
;  SEQUENCE DESCRIPTION: SEQ ID NO: 9 :
US-09-840-787-9

```

```

Query Match          100.0%;  Score 2764;  DB 3;  Length 531;
Best Local Similarity 100.0%;  Pred. No. 1e-166;
Matches 531;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1  MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfvQ  60
|
Db      1  MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfvQ  60
|
Qy     61  YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
|
Db     61  YKATSLEKQHKHDLLETPDLGVTIDLINPDtyRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
|
Qy    121  SQQHAKVVPWmrKTEYISTEFNryGISNEKPEVKIGSVKQqFTEEEIYKDRDSQITAIE 180
|
Db    121  SQQHAKVVPWmrKTEYISTEFNryGISNEKPEVKIGSVKQqFTEEEIYKDRDSQITAIE 180
|
Qy    181  KTFEDAQKSISQHYskPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
|
Db    181  KTFEDAQKSISQHYskPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
|
Qy    241  MSQAMIRGMMDEEGNQFvAYFLPVEETLKKRKRdQEEEMDYAPDDVYDYKIAREYNWNV 300

```

Db	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWNVK	300
Qy	301	NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Db	301	NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Qy	361	LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	420
Db	361	LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER	420
Qy	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN	480
Db	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN	480
Qy	481	DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531
Db	481	DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531

# RESULT 2

US-10-721-553-2

```

; Sequence 2, Application US/10721553
; Publication No. US20050032079A1
; GENERAL INFORMATION:
; APPLICANT: Batra, Surinder K.
; APPLICANT: Hollingsworth, Michael A.
; APPLICANT: University of Nebraska Board of Regents
; TITLE OF INVENTION: Novel Gene That is Amplified and
; TITLE OF INVENTION: Overexpressed in Cancer and Methods of Use Thereof
; FILE REFERENCE: UNMC63121
; CURRENT APPLICATION NUMBER: US/10/721,553
; CURRENT FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: US/09/647,143
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US99/06633
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/079,649
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-721-553-2

```

Query Match		100.0%;	Score 2764;	DB 5;	Length 531;
Best Local Similarity		100.0%;	Pred. No. 1e-166;		
Matches	531;	Conservative	0;	Mismatches	0;
		Indels	0;	Gaps	0;
Qy	1	MAPTQQTQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPFDQNRQVQ	60		
Db	1	MAPTQQTQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPFDQNRQVQ	60		
Qy	61	YKATSLEKQHKHDLTLPDLGVTIDLPNPDYRIDPNVLLDPADEKLLLEEEIQAPTSSKR	120		

Db	61	YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLLEEEIQAPTSSKR	120
Qy	121	SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE	180
Db	121	SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE	180
Qy	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	240
Db	181	KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM	240
Qy	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK	300
Db	241	MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK	300
Qy	301	NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Db	301	NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE	360
Qy	361	LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGSEDEHSGSESER	420
Db	361	LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGSEDEHSGSESER	420
Qy	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN	480
Db	421	EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN	480
Qy	481	DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531
Db	481	DSDSGSNGGGQSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD	531

RESULT 3

US-10-450-763-50041

; Sequence 50041, Application US/10450763

; Publication No. US20050196754A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 790CIP3/US

; CURRENT APPLICATION NUMBER: US/10/450,763

; CURRENT FILING DATE: 2003-06-11

; PRIOR APPLICATION NUMBER: PCT/US01/08631

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 09/540,217

; PRIOR FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: 09/649,167

; PRIOR FILING DATE: 2000-08-23

; NUMBER OF SEQ ID NOS: 60736

; SOFTWARE: Custom

; SEQ ID NO 50041

; LENGTH: 553

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: DOMAIN

; LOCATION: (377)..(428)

```
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by
eMATRIX,
; OTHER INFORMATION: accession number BL00412D, p-value=9.633e-09, raw score
of 16.54
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(553)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-50041
```

```
Query Match          96.2%; Score 2658.5; DB 5; Length 553;
Best Local Similarity 95.0%; Pred. No. 5e-160;
Matches 515; Conservative 6; Mismatches 10; Indels 11; Gaps 1;
```

```
Qy      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRFVQ 60
      |||
Db      12 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRFVQ 71

Qy      61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
      |||
Db      72 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 131

Qy      121 SQQHAKVVPWMrKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180
      |||
Db      132 SQQHAKVVPWMrKTEYISTEFNRYCIFHEKPEVKKWGSVKQQFTEEEIYKDRDSQITAIE 191

Qy      181 KTFEDAQS-----ISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAP 229
      |||
Db      192 KTFEDAQKSVIEGLGWGEARISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAP 251

Qy      230 KDTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDY 289
      |||
Db      252 KDTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDY 311

Qy      290 KIAREYNWNVKNKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALL 349
      |||
Db      312 KIAREYNWNVKNKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALL 371

Qy      350 VVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGS 409
      |||
Db      372 VVKHRDMNEKELEAQETRKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSEKEGS 431

Qy      410 EDEHSGSESEREEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDE 469
      |||
Db      432 EDEHSGSESEREEGDRHEASDKSGSGQDDSSDYXARAARDKEEIFGSDADSEDDADSDDE 491

Qy      470 DRGQAQGGSDNDSDSGSGNGGGQSRSHRSASPFPSPGSEHSAQEDGSEAAASDSSEADSD 529
      |||
Db      492 DRGQAQGGSDNDSDSGRNGGGQRTSRSHRSASPFPSPGSEHSAQENGSEAAASDSSEADSD 551

Qy      530 SD 531
      ||
Db      552 SD 553
```

```
RESULT 4
US-09-986-480-410
```

```
; Sequence 410, Application US/09986480
; Publication No. US20030027999A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/12788
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 410
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (405)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-986-480-410
```

```
Query Match          89.1%; Score 2464; DB 3; Length 473;
Best Local Similarity 99.8%; Pred. No. 8.5e-148;
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPFDQNRVQ 60
|
Db      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCVKYCNLSLPDIPFDPKFITYPFDQNRVQ 60

Qy     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
|
Db     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120

Qy    121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180
|
Db    121 SQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180

Qy    181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
|
Db    181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240

Qy    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300
|
Db    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy    301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360
|
Db    301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360

Qy    361 LEAQEARKAQLLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSEKEGSEDEHSGSESER 420
|
Db    361 LEAQEARKAQLLENHEPEEEEEEMETEEKEAGGSDEEQEGSSSXKEGSEDEHSGSESER 420

Qy    421 EEGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQ 473
```

|||||  
Db 421 EEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGQ 473

RESULT 5

US-10-863-332-410

; Sequence 410, Application US/10863332

; Publication No. US20050064458A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: 143 Human Secreted Proteins

; FILE REFERENCE: PS500P1

; CURRENT APPLICATION NUMBER: US/10/863,332

; CURRENT FILING DATE: 2004-06-09

; PRIOR APPLICATION NUMBER: US/09/986,480

; PRIOR FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: PCT/US00/12788

; PRIOR FILING DATE: 2000-05-11

; PRIOR APPLICATION NUMBER: US 60/134,068

; PRIOR FILING DATE: 1999-05-13

; NUMBER OF SEQ ID NOS: 456

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 410

; LENGTH: 473

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (405)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-863-332-410

Query Match 89.1%; Score 2464; DB 5; Length 473;

Best Local Similarity 99.8%; Pred. No. 8.5e-148;

Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRFVQ 60
        |||
Db      1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRFVQ 60

Qy     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120
        |||
Db     61 YKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120

Qy    121 SQQHAKVVPWMrKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180
        |||
Db    121 SQQHAKVVPWMrKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE 180

Qy    181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
        |||
Db    181 KTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALEM 240

Qy    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300
        |||
Db    241 MSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVK 300

Qy    301 NKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
```

```

Db          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| 360
301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360

Qy          361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESER 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGGSDEEQEKGSSSXKEGSEDEHSGSESER 420

Qy          421 EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQ 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 EEGDRDEASDKSGSGEDESEDEARAARDKEEIFGSDADSEDDADSDDEDRGQ 473

```

RESULT 6

```

US-11-097-143-4281
; Sequence 4281, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4281
; LENGTH: 538
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-4281

```

```

Query Match          45.0%; Score 1244.5; DB 6; Length 538;
Best Local Similarity 50.0%; Pred. No. 1.5e-70;
Matches 271; Conservative 66; Mismatches 172; Indels 33; Gaps 11;

```

```

Qy          1 MAPTIQTQAQREDGHRPNSHRTLPERSGVVCrvKYCNSLPDIPFDPKFITYPFDQNRfvQ 60
          | ||| | : | : : |||| : ||||| ||: |||| : |||
Db          1 MPPTINNSAVNSAAEK-RPQRQTERKSEIICrvKYGNLPLDIPFDLkFLQYPFDShRFVQ 59

```

Qy 61 YKATSLEKQHKHDLLEPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKR 120  
 | | | | : | : | | | | | | : | : | | | | | | | | | | | |  
 Db 60 YNPTSLERNFKYDVLTEHDLGVTVDLINRELYQADSMTLLDPADEKLLEEEETLTPTDSVR 119

Qy 121 SQQHAKVVPWMRKTEYISTEFNRYGISN-EKPEVKIGVSVKQQFTEEEIYKDRDSQITAI 179  
 | : | | : | | : | | | | | | : | | | | : | : | | : | | | | |  
 Db 120 SRQHSRTVSWLRKSEYISTEQTRFQPQNLENIEAKVGYNVKKSLREETLYLDREAQIKAI 179

Qy 180 EKTTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGAAALE 239  
 | | | | | : | : | | | | | | | | : | | | | | | | | : | | |  
 Db 180 EKTFSDTKSEITKHYSKPNVVPVEVLPIFPDFTNWKFPDCAQVIFDSDPAPAGKNVPAQLE 239

Qy 240 MMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNV 299  
 | | | | | : | | | | | | | | : | : | : | | | | | | | | | |  
 Db 240 EMSQAMIRGVMDESGEQFVAYFLPTEQTLEKRRTDFINGELYKEEEYKYKIAREYNWNV 299

Qy 300 KNKASKGYEENYFFIFREGDGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEK 359  
 | | | | | | | | : | : | | | | | | | | : | | | | | | | | : |  
 Db 300 KTKASKGYEENYFFVMRQ-DGIYYNELETRVRLNKRVRKVG-QQPNNTKL VVKHRPLDSM 357

Qy 360 ELEAQEARKAQLLENHEPEEE-----EEEEM-----ETEE-----KEAGGSD----- 395  
 | | | | : | | | | | | | | : | | | : | | : | | : | | |  
 Db 358 EHRMQRYSRERQLEVPGEIIIIEVREEEQMQIIIGETEKTSEDAAGQAASGADSPAQV 417

Qy 396 --EEQEKGSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEEI 453  
 : | : | : | | | | | | : : : | | | : | : :  
 Db 418 ARDRQSRSRSTRSGS-SSGSGSGSGSRASSRSKSGSRSGSGSRSTNSPAGSQKSGSR- 475

Qy 454 FGSDADSEDDADSDDEDGQAQGGSDNDSDSGS-NGGGQSRSRSHRSASPFPSPGSEHSAQ 512  
 | : | : | | : : | : | | | : | | | | | | | | | :  
 Db 476 SRSVSRSRSRKSGSRSRSRSRKSGSRSRSGSGSRSPSRSRSGSPSGSGSSSGSA 535

Qy 513 ED 514  
 |  
 Db 536 SD 537

# RESULT 7

US-10-450-763-50040

; Sequence 50040, Application US/10450763

; Publication No. US20050196754A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 790CIP3/US

; CURRENT APPLICATION NUMBER: US/10/450,763

; CURRENT FILING DATE: 2003-06-11

; PRIOR APPLICATION NUMBER: PCT/US01/08631

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 09/540,217

; PRIOR FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: 09/649,167

; PRIOR FILING DATE: 2000-08-23

; NUMBER OF SEQ ID NOS: 60736

; SOFTWARE: Custom

; SEQ ID NO 50040

; LENGTH: 133

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-450-763-50040

Query Match 22.5%; Score 622; DB 5; Length 133;  
Best Local Similarity 64.4%; Pred. No. 7.2e-32;  
Matches 130; Conservative 0; Mismatches 2; Indels 70; Gaps 1;

```
Qy      273 RDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENYFFIFREGDGVYYNELETRVRL 332
          |||||||||||||||||||||||||||||||||||
Db      1  RDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGY----- 35

Qy      333 SKRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEKEAG 392
          |||||||||||||||
Db      36 -----EEEEEMETEEKEAG 50

Qy      393 GSDEEQEKGSSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEE 452
          || |||||||||||||||||||||||||||||||||||
Db      51 GSYEEQEKGSSSEKEGSEDEHSGSESEREEGDRDEASDKSGSGEDESSEDEARAARDKEE 110

Qy      453 IFGSDADSEDDADSDDEDRGQA 474
          ||||||||||| |||||
Db      111 IFGSDADSEDDADSYDEDRGQA 132
```

RESULT 8

US-10-424-599-223174

; Sequence 223174, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 223174

; LENGTH: 86

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_43557C.1.pep

US-10-424-599-223174

Query Match 16.4%; Score 452; DB 4; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5e-21;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      66 LEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKRSQQHA 125
          |||||||||||||||||||||||||||||||||||
Db      1  LEKQHKHDLLETPDLGVTIDLINPDYRIDPNVLLDPADEKLLEEEIQAPTSSKRSQQHA 60
```

Qy 126 KVPWPMRKTEYISTEFNRYGISNEKP 151  
 |||||  
 Db 61 KVPWPMRKTEYISTEFNRYGISNEKP 86

RESULT 9

US-10-424-599-213359

; Sequence 213359, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
 With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 213359

; LENGTH: 571

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_34688C.1.pep

US-10-424-599-213359

Query Match 13.6%; Score 375.5; DB 4; Length 571;  
 Best Local Similarity 27.2%; Pred. No. 1.6e-15;  
 Matches 128; Conservative 71; Mismatches 170; Indels 101; Gaps 19;

Qy 14 GHRPNSHRTLPE---ER-----SGVVCVKYCNLPDIPFDPKFITYPFQNRVQY 61  
 | | | : | | : : : : | | | | : | : : : |  
 Db 150 GSRMGERRSTPLLGAERVENRLKKPTTFLCKLKFRNELPDPSAQPKLMASKKDKDQYAKY 209  
 Qy 62 KATSLEKQHKHDLLETPDLGVTIDLINPDYRIDPNVL--LDPADKLLLEE-----IQA 114  
 | | | : | | | : : : : | | : | | : : : : | :  
 Db 210 TITSLEKMYKPKLFVEPDGLPLDLLDSVYN-PPSVRPPLAPEDKELLRDDEAVTPIKK 268  
 Qy 115 PTSSKRSQQHAKVVPWPMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEI----- 168  
 : : : | | : | : | : : | : | | | : : :  
 Db 269 DGIKRKERPTDKGVAWLVTQYISP-----LSME-----STKQSLTEKQAKELREM 314  
 Qy 169 -----YKDRDSQITAEKTFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCA 219  
 | : | | | : | | | | : : | | | : | | : :  
 Db 315 KGGRGILDNLNSRERQIREIEASFE-AAKSDPVHATNKDLYPVEVMPLLPDFDRYDDQFV 373  
 Qy 220 QVIFDSDP-----APKDTSGAAALE---MMSQAMIRGMMDEEGNQFVAYFLPVEETLK 269  
 | : | | | | : | : | : : | : | : | : |  
 Db 374 VAAFDNAPTADSEMHAAMDKSVRDAFESKAVMKSYPVATGSDPANPEKFLAYMVPAPGELS 433  
 Qy 270 KRKRQDEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENYFFIFREGDGVYNELETR 329  
 | | : : : | : : : : : : : | : | : | :  
 Db 434 KDIYDENEDVSY-----WIREYHWDVRGDDADD-PATFLVAFDESEARYL-PLPTK 483

```

Qy      330 VRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEMETEEK 389
      : | | : | | | | : | | | | | | | | | | | | | | | | |
Db      484 LVLRRKKRAKEG-RSG-----DEVEQCPVPARVTVRRRSSVAAIERK 523

Qy      390 EAGGSDEEQEKGSSSEK-----EGSEDEHSGS---ESEREEDRDEASD 430
      :: | | | : | | : | | : | | : | | | | | | | | | |
Db      524 DSG--VYTSSKGNSSKRGGLEMDGLEDQHRGAPHQDNYQSSGAEDYMSD 571

```

RESULT 10

US-10-437-963-116147

; Sequence 116147, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules

Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 116147

; LENGTH: 644

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_19676C.1.pep

US-10-437-963-116147

Query Match 12.2%; Score 337.5; DB 4; Length 644;

Best Local Similarity 25.3%; Pred. No. 4.8e-13;

Matches 109; Conservative 82; Mismatches 156; Indels 83; Gaps 17;

```

Qy      17 PNSHR---TLPERSGVVCRVKYCNSLPDIPDPKFITYPFDQNRQVQYKATSLEKQHKHD 73
      | | : | | : : : | : | | | | | | | | | | | | | |
Db      228 PNAERFENRLKKPTTFLCKHKFRNELPDPSSQLKWLPLNKDKDRYTKYRITSLEKNYIPK 287

Qy      74 LLTEPDLGVTIDLINPDYRIDPNVL-----LD 101
      :: | | | : | | : | : : : : : : : : : : : : :
Db      288 MIVPEDLGIPLDLLDMSVYKYELSQIYVFLAAILIRRFLLAYLTYAQTIFSTPPVQPPMA 347

Qy      102 PADEKLL-EEEIQAPTSS----KRSQQHAKVVPWMRKTEYI---STEFNRYGI----SNE 149
      | | | : | | : | | : | : | | | | | | | | | |
Db      348 PEDEELLRDDEVLTVPVKDGIRKKERPTDKGMSWLVKTYISPLSTDAAKMSITEKQAKE 407

Qy      150 KPEVKIGVSVKQQFTEEEIYKDRDSQITAIEKTFEDAQKSISQHYSKPRVTPVEVMPVFP 209
      : | : | : | | | | | | | | | | | | | | | |
Db      408 RRESREG---RNTFLEN--INDREKQIKAIEDSFR-AAKSRPVHQTKRGMEAEWVLP LLP 461

```

```

Qy      210 DFKMWINPCAQVIFDSDPAPKDTSGAAALE-----MMSQAMIRGMMDEEGNQFV 258
      || : : | || || | : || : | : : | :
Db      462 DFDYDDQFVMVNFDDGPT-ADSEQYNKLSERDECESRAVMKSFLVNGSDPAKQEKFL 520

Qy      259 AYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENYFFIFREG 318
      || : | | | : | : | : || : | | : | :
Db      521 AYMVPSPHELKDLDDDETEDIQYS-----WLREYHWEVRGD-DKDDPTTYLVTF-DD 570

Qy      319 DGVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLN---HE 375
      || | | : : | : : || | : || : | : : : :
Db      571 DGAHYLPPLPTKLVLQKKKAKEG-RSGDE----IEHFPVPSRITENLKRQRSSVDDDLVDH 625

Qy      376 PEEEEEEEME 385
      | : | :
Db      626 PKHSRVEDMD 635

```

RESULT 11

US-10-450-763-49771

; Sequence 49771, Application US/10450763

; Publication No. US20050196754A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 790CIP3/US

; CURRENT APPLICATION NUMBER: US/10/450,763

; CURRENT FILING DATE: 2003-06-11

; PRIOR APPLICATION NUMBER: PCT/US01/08631

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 09/540,217

; PRIOR FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: 09/649,167

; PRIOR FILING DATE: 2000-08-23

; NUMBER OF SEQ ID NOS: 60736

; SOFTWARE: Custom

; SEQ ID NO 49771

; LENGTH: 475

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: DOMAIN

; LOCATION: (391)..(429)

; OTHER INFORMATION: ELEMENT TRANSPOSABLE INSERTION PROTEIN TRANSPOSITION DNA

; OTHER INFORMATION: domain identified by eMATRIX, accession number PD02455A,

p-value=

; OTHER INFORMATION: 1.450e-25, raw score of 25.65

; FEATURE:

; NAME/KEY: DOMAIN

; LOCATION: (68)..(113)

; OTHER INFORMATION: Immunoglobulin domain identified by PFam, accession name

ig,

; OTHER INFORMATION: E-value=0.099, PFam score of 10.6

US-10-450-763-49771

Query Match 10.2%; Score 283; DB 5; Length 475;

Best Local Similarity 87.7%; Pred. No. 9.4e-10;

Matches 57; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

Qy      328 TRVRLSKRRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEEEMETE 387
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      3  SRVRLSKRRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEEIROP 62

Qy      388 EKEAG 392
          |:|
Db      63 RKKLG 67

```

RESULT 12

US-10-425-115-202470

; Sequence 202470, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

: APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

With

; TITLE OF INVENTION: Plants

; FILE REFERENCE: 38-21(53222)B

: CURRENT APPLICATION NUMBER: US/10/425,115

: CURRENT FILING DATE: 2003-04-28

NUMBER OF SEO ID NOS: 369326

; SEQ ID NO 202470

; LENGTH: 286

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

```
; NAME/KEY: unsure
```

; LOCATION: (1) .. (286)

; OTHER INFORMATION: unsure at all Xaa locations

FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577\_116243C.1.pep

US-10-425-115-202470

Query Match 8.6%; Score 238.5; DB 4; Length 286;

Best Local Similarity 28.0%; Pred. No. 3.4e-07;

Matches 76; Conservative 49; Mismatches 113; Indels 33; Gaps 9;

Qy 30 VCRVKYCNSLPDIPFPDPKFITYPFDQNRFVQYKATSLEKQHKHDLLETPELGVITIDLINP 89  
:: :: | ||| ::: ::: :: :||| : :: ||| :||::

Dd 23 LCKHKFERNELPDPSAQLKWPLNKKDKDRYTKYRISSLEKNYLPKMIVPEDLGIPDLLDM 82

Qy 90 DTYRIDPNVL-LDPADEKLL-EEEIQAPTS----SKRSQQHAKVVPWMRKTEYISTEFNR 143  
| | | | | : : | | : : | : | : | : | : |  
Db 83 TVYNPPAAQLPLAPEDEELLRDDEVLTVPVKPEGIRKKERPTDKGMSWLVKTYQYISP---- 138

```

Qy      144 YGISNEKPEVKIGVSVKQQTTEE-----EIYKDRDSQITAIEKTFEDAQKSISQHYs 195
          :| : :: |      :: |      | || :| || ::| : || | :
Db      139 --LSTDAAKMSITEKOAKERRESGEGRDNVLENLNDROKRIKAIAESFK-AAKSRPVHQT 195

```

QY 196 KPRVTPVEVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEMM-----SQA----- 244  
| : | |:| || : :| | || | | : || ||

Db 196 KRGMEPEFVLPLVPDFDRYNDPFVMVNFDDGPT-ADSEQYNKLSRVRDECESQAXDESF 254

Qy 245 MIRGMMDEEGNQFVAYFLPVEETLKKRKRDO 275  
: | : : || | | |:  
Db 255 XVSGSDPAKXREILAYMXSSPHELVKDLDDDE 285

RESULT 13

US-10-282-122A-70437

; Sequence 70437, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 70437  
; LENGTH: 1633  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-70437

Query Match

8.6%; Score 237.5; DB 4; Length 1633;

Best Local Similarity 21.2%; Pred. No. 3.1e-06;  
Matches 131; Conservative 85; Mismatches 238; Indels 165; Gaps 25;

```

Qy      48 FITYFPDQNRVFQYKATSLEKQHKHDLLEPDLGVTIDLINPDYRIDP-----NVLLDP 102
      :|      || :|      |      :: | |      |:: || |      |      | :|
Db      707 YVTLKDSNNRELQRVTTDQSGHYQFDNLQNGT--YTVEFAIPDNYTPSPANNSTNDAIDS 764

Qy      103 ADEKLLEEEIQAPTSSKRSQQHAKV-----VPWMRKTEYISTEFNRYGI--SNEKPEVK 154
      |:      ::      :      :      : |      |      :|:      :|: ||      |||
Db      765 DGERDGRKVVVAKGTINNADNMTVDTGfYLTpKYNVGDYVWEDTNKDGIQDDNEKGISN 824

Qy      155 IGVSVKQ-----QFTEEEIY-----KDRDSQ 175
      : |::|      :|      | |      : :||
Db      825 VKVTLKNKNGDTIGTTTTDSNGKYEFTGLENGDYTIEFETPEGYTPTKQNSGSDEGKDSN 884

Qy      176 ITAIEKTFEDA-QKSISQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFSDP----- 227
      |      | :||      |:|      : ||      |: :| :      | |
Db      885 GTKTTTVTKDADNKTIIDSGFYKPIYN-----LGDY-VWEDTNKDGIQDDSEKGISGVK 936

Qy      228 -APKDTSGAA-----ALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEMDY- 281
      || :| |      :      :|: | | : | : | |      | : :|
Db      937 VTLKDKNGNAIGTTTTDASGHYQFKGL--ENGsyTVEFETPSGYTPTKANSGQDITVDSN 994

Qy      282 -----APDDVYD---YKIAR---EYNWNVKNK-----ASKGY----- 307
      | :      | || :      :| | ||      ||
Db      995 GITTTGIINGADNLTIDSGFYKTPKYSVG DYVWEDTNKDG IQDDNEKGISGVK VTLKDEK 1054

Qy      308 -----EENYFFIFREGD-GVYYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMN 357
      :|| : |      | | |      | ::: | :|      : :
Db      1055 GNIISTTTTIDENGKYQFDNLDSGNYIIHFKEKPEGMTQT TANSG-----NDD 1100

Qy      358 EKELEAQEARKAQLNH-----EPEEEEEEMETEEKEAGGSD 395
      ||: : :: |      : :|      : : : : : : : ||
Db      1101 EKDADGEDVR-VTITDHDDFSIDNGYFDDSDSDSDADSDSDSDSDADSDSDADSDSN 1159

Qy      396 EEQEKGSSEKEGSEDEHSGSESERE-EGDRDEASDK-SGSGEDESSEDEARAARDKEEI 453
      : : | |: :      | | |:|: : : | | || | | | | : : : : | :
Db      1160 SDSDSDSDSDSDSDSDSDSDSDSDSDADSDSDADSDSDSDSDSDSDSDSDSDSDSDSDSD 1219

Qy      454 FGSDADSEDDADSDDEDRGQAQGGSDNDSDSGSNGGGQSRSHRSASPFPSGSEHSAQE 513
      |||||: ||||| :      :      ||:|||| |:      | | | | | | | : :
Db      1220 SDSDADSDSDADSDSDADSDSDADSDSDSDSDSDAD---SDSDSDSDSDADSDSDSDSDS 1276

Qy      514 DGSEAAASDS-SEADSDSD 531
      |      : ||| |:|||||
Db      1277 DADSDSDSDSDSDADSDSD 1295

```

# RESULT 14

US-10-615-383-4

; Sequence 4, Application US/10615383

; Publication No. US20040038327A1

; GENERAL INFORMATION:

; APPLICANT: FOSTER, Timothy

; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCI

; FILE REFERENCE: P06335US03/BAS

; CURRENT APPLICATION NUMBER: US/10/615,383  
; CURRENT FILING DATE: 2003-07-09  
; PRIOR APPLICATION NUMBER: 09/386,962  
; PRIOR FILING DATE: 1999-08-31  
; PRIOR APPLICATION NUMBER: 60/098,443  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/117,119  
; PRIOR FILING DATE: 1999-01-25  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1742  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-615-383-4

Query Match 8.5%; Score 234.5; DB 4; Length 1742;  
Best Local Similarity 21.0%; Pred. No. 5.2e-06;  
Matches 127; Conservative 88; Mismatches 240; Indels 149; Gaps 24;

Qy 48 FITYFPDQNRVQYKATSLEKQHKHDLLETPDLGVTIDLINPDYRIDP-----NVLLDP 102  
: : | | : | | : : | | | : : | | : |  
Db 716 YVTLKDSNNRELQRVTTDQSGHYQFDNLQNGT--YTVEFAIPDNYTPSPANNSTNDAIDS 773  
Qy 103 ADEKLLEEEIQAPTSSKRSQQHAKV-----VPWMRKTEYISTEFNRYGISNEKPEVKIG 156  
| : : : : : | : : : | : : : |  
Db 774 DGERDGRKVVVAKGTINNADNMTVDTG FYLTPKYNVGDYVWEDTNKDG IQDDNEKGISG 833  
Qy 157 VSV-----KQQFT--EEEIY-----KDRDSQ 175  
| | : | | | : : |  
Db 834 VKVT LKNKNGDTIGTTTTDSNGKYEFTGLENGDYTIEFETPEGYTPTKQNSGSDEGKDSN 893  
Qy 176 ITAIEKTFEDA-QKSIHQHYSKPRVTPVEVMPVFPDFKMWINPCAQVIFDSDP----- 227  
| : | : | : | : | : | : | : | : |  
Db 894 GTKTTVTVKDADNKTIDSGFYKPTYN-----LG DY-VWEDTNKDG IQDDSEKGISGVK 945  
Qy 228 -APKDTSGAA----ALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDEEEMDY- 281  
| : | : : : : : : : : : : : : : :  
Db 946 VTLKDKNGNAIGTTTTDASGHYQFKGL--ENG SYTVEFETPSGYTPTKANSQG DITVDSN 1003  
Qy 282 -----APDDVYD---YKIAR----EYNWNVKNK-----ASKGY----- 307  
| : | | : : | | |  
Db 1004 GITTTGIINGADNLTIDSGFYKTPKYSVG DYVWEDTNKDG IQDDNEKGISGVKVT LKDEK 1063  
Qy 308 -----EENYFFIFREGD-GVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMN 357  
: | : | | | : : : : : : : :  
Db 1064 GNIISTTTT DENGKYQFDNLDSGNYIIHF EKPEGMTQT TANS G-----NDD 1109  
Qy 358 EKELEAQEARKA-----QLENHEPEEEEEEEEMETEEKEAGGSDEEQEKGSSSEKEGSE 410  
| : : : : : : : : : : : : : : : :  
Db 1110 EK DADGEDVRVTITDHDDFSIDNGYFDDSDSDSDADSDSDSDSDADSDSDADSDSDA 1169  
Qy 411 DEHSGSESEERE-EGDRDEASDK-SGSGEDESSEDEARAARDKEEIFGSDADSEDDADSD 468  
| | : : : : | | | | : : : : : : : : : :  
Db 1170 DSDSDSDSDADSDSDSDSDSDSDSDADSDSDSDSDSDADSDSDSDSDSDSDSDSDSDS 1229  
Qy 469 EDRGQAQGGSDNDSGSGNGGQRRSHRSASFPFSGSEHSAQEDGSEAAASDS-SEAD 527

Qy	48	FITYPFDQNR	FVQYKATSLEKQHKHDL	LT	TEPDLGV	TIDLIN	PD	TYRIDP	----	NVLLDP	102
		::	:		::		::				:
Db	716	YVTLKDSNN	RELQRV	TTDQSGHYQ	FDNLQNGT	--	YTVEFAI	PDNYTP	SPANNST	NDAIDS	773
Qy	103	ADEKLL	EEEEIQAPT	SSKRSQQ	HAKV	-----	VPWMRK	TEYISTE	FNRYGIS	NEKPEVKIG	156
		:	::	:	:	:		:	:		:
Db	774	DGERD	GTRKVVAKGT	INNADNMT	VD	TG	FYLT	PKYNVG	DYVWED	TNKDGIQDD	NEKGISG 833
Qy	157	VSV	-----	KQQFT	--	EEEIY	-----	KDRDSQ	175		
						:				:	
Db	834	VKVT	LKNKNGDTIG	TTTTDS	NGKYEF	TGLE	NDYTI	EFET	PEGYTPTK	QNSGSDEG	KDSN 893
Qy	176	ITAIEK	TFEDA-QKSISQ	HYSKPR	VT	PVEV	MPVF	PDFK	MWINPCA	QVIFDSDP	----- 227
				:		:		:		:	
Db	894	GTKT	TVTVK	DADNKTID	SGFYKPTYN	-----	LGDY	-VWED	TNKDGIQDD	SEKGISGVK	945
Qy	228	-APKDT	SGAA----	ALEMMSQ	AMIRGMM	DDEEGN	QFVAY	FLPVEET	LKKRKR	DQEEEMDY	- 281
			:		:	:	:		:		:
Db	946	VTLK	DKNGNAIG	TTTTDAS	GHYQFKGL	--	ENG	SYTVE	FETPSGYT	PTKANSGQDIT	VDSN 1003



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2006, 09:23:34 ; Search time 26 Seconds  
(without alignments)  
898.675 Million cell updates/sec

Title: US-10-721-553-2  
Perfect score: 2764  
Sequence: 1 MAPTIQTQAQREDGHRPNSH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA New:\*

- 1: /SIDS5/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /SIDS5/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /SIDS5/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 4: /SIDS5/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 5: /SIDS5/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 6: /SIDS5/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 7: /SIDS5/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*
- 8: /SIDS5/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	226	8.2	743	6	US-10-485-517-351	Sequence 351, App
2	226	8.2	877	6	US-10-485-517-200	Sequence 200, App
3	224	8.1	265	6	US-10-799-749-78	Sequence 78, Appl
4	220.5	8.0	1253	7	US-11-185-924-18	Sequence 18, Appl
5	212.5	7.7	287	6	US-10-793-626-468	Sequence 468, App
6	206	7.5	280	6	US-10-485-517-240	Sequence 240, App
7	183	6.6	486	6	US-10-793-626-788	Sequence 788, App
8	182	6.6	1244	7	US-11-087-099-9112	Sequence 9112, App
9	181	6.5	697	6	US-10-821-234-905	Sequence 905, App

10	178	6.4	2004	6	US-10-469-469-250	Sequence 250, App
11	177	6.4	8746	7	US-11-098-686-10232	Sequence 10232, A
12	173	6.3	499	7	US-11-087-099-1159	Sequence 1159, Ap
13	171	6.2	513	7	US-11-185-924-16	Sequence 16, Appl
14	169.5	6.1	447	7	US-11-096-568A-28367	Sequence 28367, A
15	169	6.1	1758	7	US-11-087-099-9570	Sequence 9570, Ap
16	168.5	6.1	345	7	US-11-024-959-415	Sequence 415, App
17	168.5	6.1	699	7	US-11-138-642-4	Sequence 4, Appli
18	168.5	6.1	699	7	US-11-138-882-4	Sequence 4, Appli
19	168.5	6.1	699	7	US-11-138-757-5	Sequence 5, Appli
20	167	6.0	1388	6	US-10-821-234-1143	Sequence 1143, Ap
21	165	6.0	440	7	US-11-108-172-1059	Sequence 1059, Ap
22	164.5	6.0	795	6	US-10-770-726-49	Sequence 49, Appl
23	164	5.9	1750	7	US-11-087-099-12397	Sequence 12397, A
24	162.5	5.9	302	7	US-11-079-463-8516	Sequence 8516, Ap
25	162.5	5.9	414	7	US-11-089-551A-28	Sequence 28, Appl
26	162	5.9	268	7	US-11-072-512-3158	Sequence 3158, Ap
27	162	5.9	428	7	US-11-138-642-8	Sequence 8, Appli
28	162	5.9	428	7	US-11-138-882-8	Sequence 8, Appli
29	162	5.9	428	7	US-11-138-757-9	Sequence 9, Appli
30	159.5	5.8	651	7	US-11-128-660-1	Sequence 1, Appli
31	159	5.8	1085	7	US-11-087-099-11646	Sequence 11646, A
32	157.5	5.7	520	7	US-11-096-568A-16948	Sequence 16948, A
33	157	5.7	1046	7	US-11-165-819-6	Sequence 6, Appli
34	157	5.7	1046	7	US-11-165-819-26	Sequence 26, Appl
35	156	5.6	1028	7	US-11-232-440-43	Sequence 43, Appl
36	155.5	5.6	1549	7	US-11-096-568A-31271	Sequence 31271, A
37	154.5	5.6	733	7	US-11-096-568A-29646	Sequence 29646, A
38	154.5	5.6	735	7	US-11-096-568A-29645	Sequence 29645, A
39	154.5	5.6	735	7	US-11-096-568A-31569	Sequence 31569, A
40	154.5	5.6	768	7	US-11-096-568A-29644	Sequence 29644, A
41	154.5	5.6	768	7	US-11-096-568A-31568	Sequence 31568, A
42	154.5	5.6	770	7	US-11-096-568A-31567	Sequence 31567, A
43	153.5	5.6	608	7	US-11-024-959-315	Sequence 315, App
44	152.5	5.5	376	7	US-11-087-099-11305	Sequence 11305, A
45	152.5	5.5	1609	7	US-11-087-099-4181	Sequence 4181, Ap

#### ALIGNMENTS

##### RESULT 1

US-10-485-517-351

```
; Sequence 351, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
```

; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 351  
; LENGTH: 743  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-351

Query Match 8.2%; Score 226; DB 6; Length 743;  
Best Local Similarity 22.6%; Pred. No. 1.1e-06;  
Matches 125; Conservative 75; Mismatches 208; Indels 146; Gaps 23;

```
Qy      33 VKYCNSLPDIPF-DPKFITYPFDQNRVQYKATSLEKQHKHDLLEPDLGVTIDLIN--- 88
      | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db     137 VDYSNSNNTMPIADIK-----STNGDVVAKAT-----YDILTKTYTFVFTDYVNNKE 183

Qy      89 -----PDTYRIDPNVLLDPADEKLLEE-----EIQAPTSSKRS 121
      | : | | : : | | | | : | | | | | | | | | | | | | | | |
Db     184 NINGQFSLPLFTDRAKAPKSGTYDANINI--ADEMFNNKITYNYSSPIAGIDKPNGANIS 241

Qy     122 QQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKDRDSQITAIE- 180
      | | | : | | | | | | | | | | | | | | | | | | | | | |
Db     242 SQIIGVDTASGQNTYKQTVF-----VNPQQRVLGNTWVYIKGYQDKI-EESSGKVSATDT 295

Qy     181 --KTFE--DAQKSISQHYSKPRVTPV-EVMPVFPDFKMWINPCAQVIFDSDPAPKDTSGA 235
      : | | | | : | | : : | | | : : | | | | | | | | | |
Db     296 KLRIFEVNDTSKLSDSYADPNDSNLKEVTDQFKNRIYYEHPNVASIKFGD----- 346

Qy     236 AALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQEEEMDYAPDDVYDYKIAREY 295
      : : : | | | | | | | | | | | | | | | | | | | | | |
Db     347 --ITKTYVVLVEGHYDNTG-----KNLKTQVIQENVDPVTNRDYSI---F 386

Qy     296 NWNVKNKASKGYEENYFFIFREGDGVYVYNELETRVRLSKRRAKAGVQSGTNALLVVKHRD 355
      | | : | | | | | | | | | | | | | | | | | | | | | |
Db     387 GWNEN-----VVRYG-----GGSADGDSA----- 406

Qy     356 MNEKE-----LEAQEARKAQLE-----NHEPEEEEEEMETEEKEAGGSDEEQEK 400
      : | | : : : : : | | | | : : : : | | | | | | | | | |
Db     407 VNPKDPTPGPPVDPEPSPDPEPEPTPDPEPSPDPEPEPSPDPDPDSDSDSGSDSDSGS 466

Qy     401 GSSSEKEGSEDEHSGSESERE-EGDRDEASD-KSGSGEDESEDEARAARDKEEIFGSDA 458
      | | | : | | | | : : | | | | | | | | | | | | | | | |
Db     467 DSDSESDSDSDSDSDSDSDSDSDSESDSDSESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 526

Qy     459 DSEDDADSDDEDRGQAQGGSDNDSDSGSGGGQSRSHSRASPFPSGSEHSAQEDGSEA 518
      | | : | | | | : : | | | | | | | | | | | | | | | |
Db     527 DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 585

Qy     519 AASDS-SEADSDSD 531
      : | | | | : | | | | |
Db     586 SDSDSDSDSDSDSDSD 599
```

RESULT 2  
US-10-485-517-200  
; Sequence 200, Application US/10485517

```

; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 200
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-200

```

[illegible]

### RESULT 3

US-10-799-749-78

```
; Sequence 78, Application US/10799749
; Publication No. US20060020391A1
; GENERAL INFORMATION:
; APPLICANT: Kreiswirth, Barry N
; APPLICANT: Nadich, Steven M
; TITLE OF INVENTION: System and Method for Tracking and Controlling Infections
; FILE REFERENCE: 19124.0002
; CURRENT APPLICATION NUMBER: US/10/799,749
; CURRENT FILING DATE: 2004-03-15
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-799-749-78
```

Query Match 8.1%; Score 224; DB 6; Length 265;  
Best Local Similarity 34.4%; Pred. No. 4.3e-07;  
Matches 55; Conservative 29; Mismatches 72; Indels 4; Gaps 4;

Qy	375	EPEEEEEEEEMETEEKEAGGSDEEQEKGSSEKEGSEDEHSGSESERE-EGDRDEASDK-S	432
		: : ::       :      : : :	
Db	22	EPEFSPDPDFDSDSDSDSGSDSDSGSDSDSEDSDSDSDSDSDSDSEDSDSEDSDSDS	81
Qy	433	GSGEDESSSEDEARAARDKEEIFFGSADASEDDADSDEEDRGQAQGGSNDNSD SGSNNGGGQR	492
		: :: :   :   :  : ::     .::   :     :	
Db	82	DSDSDSDSDSDSEDSDSDSDSDSDSDSDSEDSDSDSEDSDSEDSDSDSDSDSDSDSD-	140
Qy	493	SRSHSRNASPFPPSGSEHSAQEDGSEA A ASDS-SEADSDSD	531
		: ::   :       :	
Db	141	SDSDSDSDSDSDSDSDSDSEDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD	180

### RESULT 4

US-11-185-924-18

```
; Sequence 18, Application US/11185924
; Publication No. US20060078945A1
; GENERAL INFORMATION:
; APPLICANT: Fisher et al., Larry
; TITLE OF INVENTION: Complex Formed by Small Integrin-Binding Ligand,
; TITLE OF INVENTION: N-Linked Glycoproteins (SIBLINGS) and Factor H
; FILE REFERENCE: 4239-61301-02
; CURRENT APPLICATION NUMBER: US/11/185,924
; CURRENT FILING DATE: 2005-07-19
```

```
; PRIOR APPLICATION NUMBER: 09/958,617
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: PCT/US00/09349
; PRIOR FILING DATE: 2000-04-09
; PRIOR APPLICATION NUMBER: 60/128,468
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 1253
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-185-924-18
```

## RESULT 5

```
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 468
;   LENGTH: 287
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-468
```

US-10-485-517-240

## RESULT 7

[illegible]

# RESULT 8

US-11-087-099-9112

; Sequence 9112, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 9112  
; LENGTH: 1244  
; TYPE: PRT  
; ORGANISM: Magnetococcus sp. MC-1  
US-11-087-099-9112

Query Match 6.6%; Score 182; DB 7; Length 1244;  
Best Local Similarity 22.9%; Pred. No. 0.0012;  
Matches 80; Conservative 49; Mismatches 150; Indels 70; Gaps 9;

Qy 212 KMWINPCAQVIFDSDPAPKDTSGAAA----LEMSQAMIRGMMDEEGNQFVAYFLPVEE- 266  
|: : | : | : | | | : | | | : : | : | |  
Db 60 KLKCSQCHHIFQAPPEPKSAQPPASEQPGLEDESTAQDNDTESRDYAEFAFEESPLEEV 119  
Qy 267 -----TLK---KRKRDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENY 311  
| | : | | : | : | |  
Db 120 DLDEIEKLTAQATLDMALEATRDKRQEPSFDEDTQVD----- 156  
Qy 312 FFIFREGDGVYYNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQL 371  
| | | | : | : : | : : | : | | | : | :  
Db 157 -----EDAVIEPSLEEVDVDQMIQAATALPTEPEAASEAEEEELEAEEEELEAEEEEPEAEE 210  
Qy 372 E---NHEPEEEEE---EEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESEREEDG 424  
| | | | | | | : | : | : | : | : | : | :  
Db 211 EPEAEEEEPEAEEEEPEAEEEELEAEEEEPEAEEEEPEAEEEESEAESEAESEAESEAESEVEE 270  
Qy 425 RDEADKSGSGEDESEDEA--RAARDKEEIFGSDADSEDDADSDDEDGQAQGGSDNDS 482  
| : : | : | | | : | | : : | : : | : | : :  
Db 271 APEVEEELELEEEAEEEEPEAEEEAPEAEEEAPEVEEEPEVEEELELEEEAEEEESEAE- 329  
Qy 483 DSGSNGGGQRSRSHRSASPFPSGSEHSAQEDGSEAAASDSSEADSDSD 531  
: | : : | | | : | | | | : :  
Db 330 -----EESEAESEEA---PEAEEEAPEAEEEAPEAEEEAPEAE 367

# RESULT 9

US-10-821-234-905

; Sequence 905, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom

```
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 905
; LENGTH: 697
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-905
```

```
Query Match          6.5%; Score 181; DB 6; Length 697;
Best Local Similarity 21.9%; Pred. No. 0.00066;
Matches 110; Conservative 68; Mismatches 184; Indels 140; Gaps 21;
```

```
Qy      108 LEEEIQAPTSSKRSQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQF---- 163
      || | : | : : || | : | : : | | : | : : :
Db      83 LEGEARTPLAIPHT-----PWGRRPEEEAEDSGGPGEDRETGLKTSSSLPEAWGLLD 135

Qy      164 TEEEIYKDRDSQITAIEK----TFEDAQKSISQHYSKPRVTPVEVMPVFPDFKMWINPCA 219
      :: : | : | : | : : | | | : | : | :
Db      136 DDDGMYGEREA--TSVPRGQGSQFADGQRA-----PLSPSLLI----- 171

Qy      220 QVIFDSDPAPKDTSGAAALEMMSQAMIRGMMDEEGNQFVAY-----FLPV-----EET 267
      : : || | : : | | : : || : | | |
Db      172 RTLQGSCKNPGE-----EKAEEEGVAEEGVNKFSPPSHRECCPAVEEEDDEEA 221

Qy      268 LKKRKRQDE-----EEMDYAPDD--VYDYKIAREYNWNVKNKAS--K 305
      : || | : | : | | : | : : : : | :
Db      222 VKKEAHTSTTSALSPGSKPSTWVSCPGEEENQATEDKRTERSKGARKTSVSPRSSGSDPR 281

Qy      306 GYEENYFFIFREGDGVYINELETRVRLSKRRAKAGVQSGTNALLV-----VKHRDMNEK 359
      : | : | : | : | | | | | : | |
Db      282 SWE-----YRSGEASEEKEEKAHKETGKGEAAPGPQSSAPAQRPQLKSWWCQPSDEEEG 335

Qy      360 ELEAQEARKAQLNH-----EPEEEEEEMETEEKEAGGSDEEQE 399
      | : | | : | | | | | | | : : | : | |
Db      336 EVKALGAAEKDGEAECPPCIPPPSAFLKAWVYWPGEDTEEEDEEEDESDSGSDEEEGE 395

Qy      400 KGSSSEKEG-----SEDEHSGSESEREEDRDEASDKSGSGEDESSEDEARAARDKEEI 453
      : || | : | : | | | | : | | | | | :
Db      396 AEASSSTPATGVFLKSWVYQPGEDTEEEE---DESD-TGSAEDE-REAETSASTPPASA 450

Qy      454 F-----GSDADSEDDADSDDEDRGQAQGGSDNDSDSGSNGGGQRSRSHRSASPFPS 505
      | | : | : | | | : : : | : | | |
Db      451 FLKAWVYRPGEDTEEEEDVDSEDKEDDSEAALGEAESDPHPSPDQRAHFRGWGYRP- 509

Qy      506 GSEHSAQEDGSEAAASDSSEAD 527
      : | | | | | :
Db      510 -----GKETEEEEAAEDWGEAE 526
```

```
RESULT 10
US-10-469-469-250
```

```
; Sequence 250, Application US/10469469
; Publication No. US20060079493A1
; GENERAL INFORMATION:
; APPLICANT: FRITZ, LAWRENCE C.
; APPLICANT: BURROWS, FRANCIS J.
; TITLE OF INVENTION: METHODS FOR TREATING GENETICALLY-DEFINED PROLIFERATIVE
; TITLE OF INVENTION: DISORDERS WITH HSP90 INHIBITORS
; FILE REFERENCE: CON-0010-USN
; CURRENT APPLICATION NUMBER: US/10/469,469
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: PCT/US02/06518
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/272,751
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 250
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-469-469-250
```

```
Query Match          6.4%; Score 178; DB 6; Length 2004;
Best Local Similarity 21.1%; Pred. No. 0.0037;
Matches 105; Conservative 65; Mismatches 163; Indels 164; Gaps 20;
```

```
Qy      109 EEEIQAPTSS-----KRSQQHAKVVPWMRKTEYISTEFNRY 144
      ||| :| ||          ||  :| ||  || ||
Db      991 EEEPESSRSSPPILTKPTLKRKKPFLHRRRVRKRKHHNSSVV-----TETIS----- 1039

Qy      145 GISNEKPEVKIGVSVKQQFTEEEIYKDRDSQ--ITAIEKTFEDAQKSISQHYSKPRVTPV 202
      |  ||          :| :| || :  :| |||
Db      1040 ----ETTEVL-----DEPFEDSDSERPMRLEPTFE----- 1066

Qy      203 EVMPVFPDFKMWINPCAQVIFSDPAPKDTSGAAALEMMSQAMIRGMMDEEGNQFVAYFL 262
      | :  :| :  | :  :  | :  :  | :  :  |
Db      1067 -----IDEEEEEDEN-----ELFPREYFRLSSQD-----VL 1094

Qy      263 PVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNW----NVKN-----KASKGY 307
      : : |:: :|:| |  ||  |  :  :  :|||  | ||:
Db      1095 RCQSSSKRKSKDEEE--DEESDDADDTPIPKPVSLLRKRDVKNSPLEPDTSTPLKKKKGW 1152

Qy      308 EENYFFIFREGDGVYNELETR---VRLSKRRAKAGVQSGT-NALLVVKHRDMNEKELEA 363
      :  :  ::: :  |  :||:  | :  :  :  |  |:
Db      1153 PKG-----KSRKPIHWKKRPGRKPGFKLSREIMPVSTQACVIEPIVSIPKAGRKPKIQES 1207

Qy      364 QEARKAQLNHEPEE-EEEEEMETEEKEAGGSDEEQEKGSSSEKEGSEDEHSGSESEREE 422
      :| : : :  ||| :|||:| :||  :||  |  :  :|:| :|
Db      1208 EETVEPKEDMPLPEERKEEEEMQAEAAEEGEEEDAASSEVPAASPADSSNSPETETKE 1267

Qy      423 GDRDEASDKSG-SGEDESSEDEARAARDKEEIFGSDADSE----DDADSDDEDRGQAQGG 477
      : :| :|  |  |||:| :  :  |  || :|  || :||| |  :
Db      1268 PEVEEEEEKPRVSEEQRQSEEEQQELEEPEPEEEEDAAAETAQNDDHDADDEDDGHLEST 1327

Qy      478 SDND-----SDSGSNGGGQSRSS-----HSRSASPPFSGSEHSA 511
      :  :  :| :  | :||  |  |  |  |
Db      1328 KKKELEEQPTREDVKEEPGVQESFLDANMQKSREIKDKKEETELDSEEEQPSHDTSVVSE 1387
```

Qy	60	QYKAT--SLEKQHKHDLLETPDLGVITIDLINPDTRYIDP-----NVLLDPADEKLLLEE	111
Db	7070	QYETTVQKIEQKYKEKKANRHILGCTLEELQEQQEESKVAVGNFTVLLEKMREKQQKEL	7129
Qy	112	IQAPTSSKRSQQHAKVVPWMRKTEYISTEFNRYGISNEKPEVKIGVSVKQQFTEEEIYKD	171
Db	7130	QQFPSSDEDESD----VDRRKKKKVIKTKAER----NAQREQR----KHHHTPHPHYHPE	7176
Qy	172	RDS-----QITAIKETFEDAQKSISQHYSKPRVTPVEVMP-----	206
Db	7177	ESTSHLVLDTKQIVTLTPSSDQETPVQSKETATNETPIPSLPSTVKGLTLEEVTVTVLP	7236
Qy	207	-----VFPDFKMWINPC----AQVIFDSDPAPKDTSGAAALEMMSQAMIRGMMDE	252
Db	7237	QPTQLSSLLSYFSSEPLQEQPCQVEEQQVSFESSQVQASTD-----EATIDPEVQLVLDE	7291
Qy	253	EGNQFVAYFLPVEETLKKRKRQDEEE--MDYAPDDVYDYKIAREYNWNVKN-----KASK	305
Db	7292	YSSKVAC---LQQEMDKKLQEIEEKGTDSSASSDT-----EWSWPKKDMPREIKTLK	7340
Qy	306	GYEENYFFIFREG-----DGVYYNELETRVRLSKR----	335
Db	7341	GSDS-----EGKDOOEVPKIPSAHAHSLSSMAESEDVGAVSHIEKKKKRKHKKHRSO	7393

Qy 336 -----RAKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEE----- 383  
 ||| : : | : : : : || | || : | : | : : ||  
 Db 7394 REKISRAKRALMAQYFAKVTLGQECSEKVSEIQEQKKQKCEIEEKQRKEREFFDQHQA 7453

Qy 384 -----METEEKEAGG-----SDEEQEKGSSSEKEGSEDEHSGSESEREEGDRDEAS 429  
 : : : | | | : : : : : : : | | : : | | : |  
 Db 7454 ARNALLAKQQQELIGVAPEEVGPLQDKHRQEQQAQSRQWGLNLHSLIKQQRKE--RGLSS 7511

Qy 430 DKSGSGEDESEDEARAARDKEE-----IFGSDADSEDDADSDDEDRGQA 474  
 | | | | | : | : : : : | : : : | | | |  
 Db 7512 SSSSSSDEDIIDEGSQSDDQEDSKSLSPISPPSPVSGADSGCIGGASSSDTDATMK 7571

Qy 475 QGG-----SDNDSDSGSN-----GGGQRSRS-HSRSASPPF 504  
 | | : : | | | | | : | | |  
 Db 7572 SDGHKSPEVPVSSDKKEETGGNQSSKVTTYLLSVFTGKGGTAAGAQSSSSEHTGSKRQQP 7631

Qy 505 SGSEHSAQE-----DGSEAAASDSSEAD 527  
 ||| : : : : : | : | : :  
 Db 7632 SGSDQTSKSSRQGPSTPFEGQGTSGVEGASGGAGDPGDGE 7671

# RESULT 12

US-11-087-099-1159

; Sequence 1159, Application US/11087099

; Publication No. US20060041961A1

; GENERAL INFORMATION:

; APPLICANT: Abad, Mark S. et al.

; TITLE OF INVENTION: Genes and Uses for Plant Improvement

; FILE REFERENCE: 38-21(53450)B EP

; CURRENT APPLICATION NUMBER: US/11/087,099

; CURRENT FILING DATE: 2005-03-22

; NUMBER OF SEQ ID NOS: 12464

; SEQ ID NO 1159

; LENGTH: 499

; TYPE: PRT

; ORGANISM: Lycopersicon esculentum

US-11-087-099-1159

Query Match 6.3%; Score 173; DB 7; Length 499;

Best Local Similarity 18.4%; Pred. No. 0.0014;

Matches 91; Conservative 86; Mismatches 185; Indels 132; Gaps 17;

Qy 101 DPADEKLLEEEIQA-----PTSSKRSQQHAKVVPWMRKTEYISTEFNR--YGISNEKPEV 153  
 | : | | : | | : : : | : : : | | : :  
 Db 6 DVIEEVLAGTEVPAIVHGVKSTKKKKK-----LWEMEAQFMKTVLGRGSYSFFDNRRNK 60

Qy 154 KIGVSVKQQFTEEEIYKDRDSQITAIEKTFEDAQKS-----ISQHYSKPR 198  
 | : | | : : : | | : | | : | : |  
 Db 61 KKSSQLFNVFQEKPDFENCNGWSTVINRKKLPALKGSQIGIYVNLTKGSMMPHWN-PM 119

Qy 199 VTPVEV-----MPVFPDFKMWINPCAQVIFSDP- 227  
 | : : : | | | : | | : : :  
 Db 120 ATEIGIAIQGEGMVRVVCSSKSGTGCKNMRFKVEEGDVFFVPRF-----DPMAQMAFNNSF 175

Qy 228 -----APKDTSG-AAALEMMSQAMIRG-----MMDEEGNQFVAYFLP 263  
 | : : | : | : : : : : : : : :  
 Db 176 VFVGFSTTTKKHHPQYLTGKASVLRITLDRQILEASFNVGNTTMHQILEAQGDSVI---LE 232

Qy 264 VEETLKKRKRQDQEEEMDYAPDDVYDYKIAREYNWNVKNKASKGYEENYFFIFREGDGVYY 323  
 :: || ||| :: : ||: : | :: : :| :  
 Db 233 CTSCAEEEEKRLMEEEMRKEEEEAKKKEEARKAEERREKEAEEERKR-----QEEEAR 287  
 Qy 324 NELETRVRLSKRRAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPE----- 377  
 | | | | : | : : :: | : | | | | | : :  
 Db 288 EEEEIRRRQEEEEEAR-----RRQEEEEERERQEARKKQEEEEEAAQREAEQA 334  
 Qy 378 --EEEEEMETEEKEAGGSDEEQEKGSSEKEGSEDEHSGSESEREEGDRDEASDKSGSG 435  
 ||| | :| : : : | :| : | :| | :  
 Db 335 RREEEEAEKRRQEEESRREEKARRRQEEARRREEEEAAKQHEEEAER-EAEEARRIE 393  
 Qy 436 EDES-----SEDEARAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDNDS 484  
 | :| : | :| | :| : :| : :| : : : :  
 Db 394 EEEAQREAEARRIQEEEAERARRREE----EAETRRKEEEEEESRRQEEESRRSEEA 449  
 Qy 485 GSNGGGQSRSHSR 498  
 : | |  
 Db 450 AREAERERQEEAER 463

# RESULT 13

US-11-185-924-16

; Sequence 16, Application US/11185924  
 ; Publication No. US20060078945A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fisher et al., Larry  
 ; TITLE OF INVENTION: Complex Formed by Small Integrin-Binding Ligand,  
 ; TITLE OF INVENTION: N-Linked Glycoproteins (SIBLINGS) and Factor H  
 ; FILE REFERENCE: 4239-61301-02  
 ; CURRENT APPLICATION NUMBER: US/11/185,924  
 ; CURRENT FILING DATE: 2005-07-19  
 ; PRIOR APPLICATION NUMBER: 09/958,617  
 ; PRIOR FILING DATE: 2002-01-18  
 ; PRIOR APPLICATION NUMBER: PCT/US00/09349  
 ; PRIOR FILING DATE: 2000-04-09  
 ; PRIOR APPLICATION NUMBER: 60/128,468  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 16  
 ; LENGTH: 513  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-11-185-924-16

Query Match 6.2%; Score 171; DB 7; Length 513;  
 Best Local Similarity 22.4%; Pred. No. 0.0019;  
 Matches 66; Conservative 54; Mismatches 117; Indels 58; Gaps 11;

Qy 231 DTSGAAALEMMSQAMIRGMMDEEGNQFVAYFLPVEETLKKRKRQDQEEEMDYAPDDVYDYK 290  
 ::| : | : | :| : : | : :| : :| : :  
 Db 233 NSAGMKSKEGENSEQANTQDSGGSQLEH--PSRKIFRKSRISEEDDRSELDDN----- 285  
 Qy 291 IAREYNWNVKNKASKGYEENYFFIFREGDGVYYNELETRVRLSKRRAGVQSGTNALLV 350  
 | : || | :| : :| :| | : |



US-11-087-099-9570

Query Match 6.1%; Score 169; DB 7; Length 1758;  
Best Local Similarity 23.2%; Pred. No. 0.011;  
Matches 116; Conservative 62; Mismatches 178; Indels 144; Gaps 27;

Qy	94	IDPNVLLDPADEKLEEEIQAPTSSKRSQQHAKVVPWMRKTEY--ISTEFNRYGISNEKPE	152
Db	878	IEGNVHVIEVDPKKAEDSLQGQVDMK---NYIMVV---KKEYEAATAYLR-----	921
Qy	153	VKIGVSVKQQFTEEEIYKDRDSQITAIEKTFEDAQKSISQHYS--KPRVTPVEVMPVFDPD	210
Db	922	--MMGVDV----TAQEV-KNEDEE-----EDGEKSKKKRRSRKRPKKKKKKKTTTGKD	967
Qy	211	FKMWINPCAQVIFD-----SDPAPKDTSGAAALEMSQ	243
Db	968	-----AQEYLDDQEQRNLEDGQDDKKDEQDDDNEKDERDDADEKDEDGQANVDEKD-	1018
Qy	244	AMIRGMMDDEEGNQFVAYFLPVEE-----TLKKRKRQDQEEEMDYAPDDVYDYKI	291
Db	1019	---KDEMSEENS-DEYNATAAEQSPQPCKSKINNAKRKRKRQKQKAEQAAKE----KA	1070
Qy	292	AREYNWNVK-----NKASKGYEENYFFIFREGDGVYNELETNRVRLSKRRAKAGVQSGT	345
Db	1071	ERE-AWEKKKQEAERKERKKREEEAQKAARKKEA---REKETREK--EAREKAAREKEE	1124
Qy	346	NAL-----LVVKHRDMNEKELEAQEARKAQLNHEPEEEEEEEEEMETEEKEAGGSDEEQ	398
Db	1125	KAMREKAAREIAARKKEAREKETREKEAREKAAREKEAREKTAREKEAREKEV--REKEA	1182
Qy	399	EKGSSSEKEGSE-DEHSGSESEREERG-DRDEASDKSGSGEDESEDEARAARDKEE----	452
Db	1183	RKKEAREKQVREKDAREKAAKEREKVAREKEAQK--ARERQEQEKEAQKAREQQEQEAR	1240
Qy	453	-----IFGSDADSEDDADSDDEDRGQAQGGSDNDSDSG-----SNGGGQRS	493
Db	1241	ERKKLDEVIVVEEKVNEDDIKQEDEVKEEA----DNNPMSSQALMPSQQQMPPGP GPGMM	1296
Qy	494	RSHRSASPFPSGSEHSAQE	513

Db 1297 LGHQ---IPPPPGLEPKPQQ 1313

Search completed: April 25, 2006, 09:26:39  
Job time : 28 secs